

**SPRACHSYNTHESE NACH REGELN
FÜR UNBESCHRÄNKTEN DEUTSCHEN TEXT**

DISSERTATION

zur
Erlangung des Grades eines
Doktor-Ingenieurs
der
Abteilung für Elektrotechnik
an der Ruhr-Universität Bochum

von

HANS-WILHELM RÜHL

Voerde/Ndrhh.

Bochum, 1984

**SPRACHSYNTHESE NACH REGELN
FÜR UNBESCHRÄNKTEN DEUTSCHEN TEXT**

DISSERTATION

zur
Erlangung des Grades eines
Doktor-Ingenieurs
der
Abteilung für Elektrotechnik
an der Ruhr-Universität Bochum

von

HANS-WILHELM RÜHL

Voerde/Ndrh.

Bochum, 1984

FÜR UNBESCHRÄNKTE DEUTSCHEN TEXT
BENUTZUNG NACH REGELN

DISSERTATION

zur
Erfassung des Grades eines
Doktor-Ingenieur
an der
Hochschule Bochum
am 16.5.1984

VON

HANS-WILHELM RÜHL

Bochum

Dissertation eingereicht am : 16.5.1984
Tag der mündlichen Prüfung : 8.6.1984
Referent : Prof. Dr.-Ing. J. Blauert
Korreferent : Prof. Dr.-Ing. W. Weber

VORWORT

Die vorliegende Arbeit entstand am Lehrstuhl für Allgemeine Elektrotechnik und Akustik der Ruhr-Universität Bochum. Dem Leiter des Lehrstuhls, Prof. Dr.-Ing. Jens Blauert danke ich herzlich für sein starkes Interesse und die intensive Förderung der Arbeit.

Meinen Kollegen Dipl.-Ing. Wolfgang Kulas und Dr.-Ing. Jürgen Schröter danke ich für viele fruchtbare Diskussionen, ebenso Herrn Prof. Dr.-Ing. Matthias Uhle, Dipl.-Ing. Peter Faubel und Frau Mariann Kugler.

Den ehemaligen Studenten Raimund Weicker, Dieter May, Michael Wörmann, Wolfgang Kulas, Uwe Schuster, Manfred Albrecht, Ulrich Winkelmann und Dietmar Dreissig danke ich für ihre Mitarbeit.

Der Deutschen Forschungsgemeinschaft danke ich dafür, daß sie mir im Rahmen eines Forschungsprojekts ermöglichte, auf dem Gebiet der Sprachsythese zu arbeiten.

Für die Übernahme des Korreferates danke ich Prof. Dr.-Ing. Wolfgang Weber.

Inhaltsverzeichnis :

1.	Kurzer Überblick über die Entwicklung der Sprachsynthese	1
2.	Ein Konzept für Vorleseautomaten	4
2.1	Anforderungen an einen Vorleseautomaten	4
2.1.1	Verständlichkeit	4
2.1.2	Natürlichkeit	6
2.1.3	Echtzeit-Betrieb	7
2.1.4	Ermüdungsfreiheit	8
2.1.5	Anpassung der Umsetzungsstruktur an den Sprechvorgang beim Menschen	9
2.1.6	Technische Einschränkungen	10
2.2	Vorlesen beim Menschen	10
2.3	Konzeptentwicklung für einen Vorleseautomaten	14
2.4	Ein Vergleich mit einem anderen Modell für Vorleseautomaten	18
2.5	Einschränkung des Arbeitsumfangs	19
2.6	Zusammenfassung	20
3.	Das Konzept der Wortumsetzung	21
3.1	Prinzipielle Möglichkeiten der Wortumsetzung	21
3.2	Ein Überblick über Verfahren zur Wortumsetzung	24
3.2.1	Umsetzung mit Listen	24
3.2.2	Umsetzung nach Regeln auf Wortebene	25
3.2.3	Umsetzung nach Regeln auf Morphemebene	27
3.3	Ein Vergleich der Umsetzungen auf Wortebene und auf Morphemebene	28
3.4	Das Konzept der Wortstrukturanalyse	30
3.5	Zusammenfassung	33

4.	Die Erkennung von Morphemgrenzen	35
4.1	Ein linguistisches Modell deutscher Wörter	35
4.2	Der Kästner'sche Ansatz zur Zerlegung von Wörtern	36
4.3	Die Klusteranalyse	38
4.3.1	Beschränkung der möglichen Morphemgrenzen	38
4.3.2	Abschätzung der Zulässigkeit der Beschränkung	40
4.3.3	Bestimmung der Morphemgrenze im Grenzbereich	44
4.4	Fehler bei falschen Zerlegungen	46
4.5	Fehler der Klusteranalyse in fremdstämmigen Wörtern	50
4.6	Die Analyse von Vokalklustern	51
4.7	Zusammenfassung	52
5.	Identifizierung von Affixen und Ausnahmen	54
5.1	Die Erkennung von Affixen	54
5.1.1	Der Aufbau eines Verifizierungseintrags	57
5.1.2	Befehle zur Verifizierung von Affixen	58
5.1.3	Die Erkennung von fremdstämmigen betonten Endungen	61
5.1.4	Fehler bei der Erkennung von Affixen	62
5.2	Die Erkennung von Endungen	64
5.3	Die Erkennung von Ausnahmen	67
5.4	Zusammenfassung	68
6.	Die Umwandlung von Wörtern	69
6.1	Die Umwandlung von Text in Lautschrift	69
6.1.1	Die Darstellung der Transkriptionsregeln	71
6.1.2	Die Phonemisierung von Vokalen	72
6.1.3	Die Phonemisierung von Konsonanten	74
6.1.4	Koartikulation	76
6.1.5	Der Knacklaut	78
6.1.6	Fehler der Transkription	78
6.2	Die Bestimmung des Wortakzents	80
6.2.1	Die Akzentuierbarkeit von Affixen	82
6.2.2	Die Festlegung des Akzentmusters	85
6.2.3	Fehler bei der Bestimmung des Wortakzents	88
6.3	Zusammenfassung	88

7.	Intonation und Synthetisatoransteuerung	90
7.1	Ein einfaches Betonungsmodell	90
7.2	Die Schnittstelle zum Phonemsynthetisator	93
7.2.1	Datenformate der Schnittstelle	93
7.2.2	Der Zusammenhang zwischen Phonemen und Datenrahmen	98
7.2.3	Das Interface des Phonemsynthetisators VS6-G	99
7.3	Zusammenfassung	103
8.	Die Textvorverarbeitung	105
8.1	Die Behandlung von Schrifttext	106
8.2	Die Behandlung von Zahlen	109
8.3	Die Behandlung von Sonderzeichen und Interpunktionen	111
8.4	Die Behandlung von Steuerzeichen	113
8.5	Zusammenfassung	114
9.	Ergebnisse	115
9.1	Überblick über die geleisteten Arbeiten	115
9.2	Realisierungen von Vorleseautomaten	117
9.3	Ausblick	118
	Literaturverzeichnis	122

Anhang

A. 1	Glossar der verwendeten linguistischen Begriffe	130
A. 2	Initiale und finale Grundkluster	138
A. 3	Falsche Zerlegungen durch die Klusteranalyse	141
A. 4	Vom Interpreter ausführbare Kommandos	144
A. 5	Zur Abtrennung von Endungen	147
A. 6	Die verwendete Lautschrift	149
A. 7	Zur Textvorverarbeitung	151

1. KURZER ÜBERBLICK ÜBER DIE ENTWICKLUNG DER SPRACHSYNTHESE

Die Idee, eine sprechende Maschine zu bauen, hat die Menschen schon seit Jahrtausenden fasziniert (Flanagan, 1972). Aber erst von Kempelen gelang es im 18. Jahrhundert, eine mechanische Apparatur zu konstruieren, mit der einzelne Laute und auch kurze Wörter durch geschickte Bedienung erzeugt werden konnten (Dudley und Tarnoczy, 1950). Der Beginn der elektronischen Sprachsynthese liegt jedoch erst in den dreißiger Jahren dieses Jahrhunderts, als Dudley 1939 seinen "Voder" auf der Weltausstellung in New York vorstellte (Dudley, Riesz und Watkins, 1939) und aus den Arbeiten zur redundanzarmen Übertragung von Sprache die ersten Vocoder entstanden (Dudley, 1936, 1939).

Während der "Voder" noch von geübten "Typistinnen" von Hand bedient wurde, gelang in den fünfziger und sechziger Jahren der Nachweis, daß es möglich ist, den Syntheseteil von Vocodern mit vom Computer erzeugten Daten anzusteuern und damit verständliche Sprache zu erzeugen (Harris, 1953; Petersen, Wang und Sivertsen, 1958; Liberman et al., 1959; Holmes, Mattingly und Shearme, 1964). In den folgenden Jahren lag der Schwerpunkt der Arbeiten in der Entwicklung von Synthetisator-Strukturen, mit denen eine möglichst natürliche Sprache erzeugt werden kann. Als besonders geeignet erwiesen sich Formant-Synthetisatoren (Flanagan, 1957, 1965; Klatt, 1980) und mit wachsender Bedeutung der Digitaltechnik LPC- bzw. PARCOR-Filter mit entsprechender Anregung (Atal und Hanauer, 1971; Itakura und Saito, 1970, 1972).

Die letztgenannten Strukturen sind besonders für Artikulations-Synthetisatoren geeignet (Coker, 1967, 1968), also für Synthesizer, deren Ansteuerparameter durch Simulation des menschlichen Artikulationsvorganges berechnet werden. Formant-Synthesizer dagegen werden häufig durch eine vorgeschaltete Steuerung zu Phonem-Synthetisatoren erweitert (Soskutý, 1977; Wolf, 1981).

Phonem-Synthetisatoren auf der Basis eines Formant-Synthetisators lassen sich mit wesentlich weniger Aufwand bauen als Artikulations-Synthetisatoren und sind seit einiger Zeit kommerziell

erhältlich (Votrax, ca. 1976; Synte, 1980), dagegen gibt es von Artikulations-Synthesizern nur Labormodelle (Coker, 1968). Beide Synthetisator Typen sind inzwischen so weit entwickelt, daß damit Sprache erzeugt werden kann, die sich von natürlicher Sprache nur wenig unterscheidet (*).

Parallel zur Entwicklung von Synthetisatoren liefen Arbeiten, deren Ziel es war, die Ansteuerung der Synthetisatoren zu vereinfachen. Dies führte zunächst zur Entwicklung der o.a. Phonem- und Artikulations-Synthesizer und dann rasch zu Synthese-Systemen, die die Ansteuerparameter für die Synthetisatoren selbst aus Text erzeugten (Coker, Umeda und Browman, 1973; Allen, 1973; Carlson und Granström, 1976), also den sog. Vorleseautomaten. Diese Arbeiten sind jedoch nicht wie die Synthesizer-Entwicklung sprachunabhängig, sondern nehmen mit zunehmender Entfernung von dem spracherzeugenden Filter immer sprachspezifischeren (im Sinne von 'language', nicht von 'speech') Charakter an. Das bedeutet, daß es zwar noch möglich ist, mit einem für amerikanisches Englisch entwickelten Phonem-Synthetisator deutsche Sprache mit verminderter Verständlichkeit zu erzeugen, indem man den deutschen Lautvorrat auf den amerikanischen abbildet, aber man kann zur Erzeugung der Laute aus Text keinen für das Englische entwickelten Algorithmus benutzen.

Die vorliegende Arbeit befaßt sich mit einem Vorleseautomaten für die deutsche Sprache. Es wird das Konzept eines Vorleseautomaten erarbeitet und es werden die Verfahren entwickelt und vorgestellt, die für einen Vorleseautomaten eine Umsetzung auf Wortebene durchführen.

Aufgrund des oben beschriebenen fortgeschrittenen Entwicklungsstadiums der Synthetisator Technik war eine der Vorgaben des Projekts, das Synthese-Filter nicht selbst zu entwickeln, sondern es wurde von vornherein auf kommerziell erhältliche Phonemsyntheti-

(*) Formant-, LPC-, PARCOR- und in beschränktem Maße auch Phonemsynthesizer werden heute sehr häufig in Ansageautomaten verwendet, also in Systemen, die bei beschränktem Wortschatz Sprache aus redundanzarm gespeicherter natürlicher Sprache erzeugen.

satoren zurückgegriffen. Desgleichen wurde bei dem Projekt nicht beabsichtigt, ein nur auf einem Großrechner lauffähiges "Labor-system" zu entwickeln, sondern die Sprachausgabe sollte auch in Zwischenversionen praktisch einsetzbar sein, z.B. in der Blinden-hilfe. Es war deshalb nicht das Ziel, eine im linguistischen Sinne vollständige und fehlerfreie Umsetzung von Text in Sprache zu entwickeln, sondern es wurden unter ingenieurmäßigen Gesichtspunkten linguistische Verfahren den praktischen Erfordernissen der Sprachsynthese angepaßt und dort, wo das nicht möglich war, geeignete Verfahren selbst entwickelt.

Teile der Arbeit befassen sich mit Verfahren, deren Grundlagen von Linguisten erarbeitet wurden. Zur Beschreibung dieser Verfahren wird auf linguistische Fachausdrücke zurückgegriffen. Um dem linguistisch nicht vorgebildeten Leser der Arbeit das Lesen zu erleichtern, werden in Anhang 1 in einem Glossar die linguistischen Fachtermini erklärt.

2. EIN KONZEPT FÜR VORLESEAUTOMATEN

In diesem Kapitel soll ein Konzept für einen Vorleseautomaten entwickelt werden. Dazu werden zunächst die ergonomischen und technischen Anforderungen an Vorleseautomaten untersucht, dann der Ablauf von Spracherzeugung und Texterkennung beim Menschen vorgestellt und zu einem Modell des Vorlesens zusammengefaßt. Aus dem menschlichen Modell wird dann das Konzept des Vorleseautomaten entwickelt.

2.1 Anforderungen an einen Vorleseautomaten

Ein idealer Vorleseautomat ist ein Gerät, das sich von einem menschlichen Sprecher nur in einem Punkt unterscheidet: er erzeugt keine "Versprecher", wie sie in menschlicher Sprache häufig vorkommen. Reale Vorleseautomaten sind von diesem Ziel noch sehr weit entfernt.

Kriterien zur Beurteilung realer Vorleseautomaten ergeben sich, wenn man den Kommunikationskanal Mensch-Maschine als Analogon zur Sprachübertragung über einen Nachrichtenkanal oder zur Wiedergabe von gespeicherter Sprache betrachtet: Verständlichkeit und Natürlichkeit.

2.1.1 Verständlichkeit

Bei analogen Systemen zur Nachrichtenübertragung wird die Verständlichkeit hauptsächlich durch die Eigenschaften des Übertragungskanals, also z.B. durch Rauschen, Begrenzungen und Verzerrungen beeinflusst. Digitale signalformcodierende Systeme mit Redundanzreduktion wie u.a. die Deltamodulation fügen weitere systemspezifische Verzerrungen hinzu, Systeme mit Redundanzreduktion durch parametrische Codierung wie das LPC-Verfahren zusätzlich noch spektrale Verschiebungen und feine zeitliche Unschärfen. Trotzdem erzeugen diese Systeme immer noch eine, wenn

auch beeinträchtigte Kopie eines Originalsignals.

Vorleseautomaten dagegen reproduzieren keine Abbilder eines Sprachsignals, sondern erzeugen ein Originalsignal. Daher treten zusätzliche verständlichkeitsmindernde Fehlerarten auf wie falsche grammatikalische Analysen ("Parsing"), Akzentuierungs- und Phonemisierungsfehler, Intonationsfehler und Fehler bei der Synthese von Segmenten (Blauert, Kulas, Rühl und Schaffert, 1984). Diese Fehler führen zu einer neuen Qualität der Verständlichkeitsminderung, denn die Sprache wirkt nicht nur qualitativ gemindert, wie Hörer es von nachrichtentechnischen Systemen gewöhnt sind, sondern sie wirkt auch dialektgefärbt.

Durch diese Dialektfärbung wird die Verständlichkeit von Vorleseautomaten zeitabhängig. So wurde z.B. für das System 'PHON-TEXT' festgestellt (Breuer et al., 1979):

"Nach einer Einhör- und Gewöhnungszeit von ca. 10 bis 15 Minuten ist eine Verständlichkeit von ca. 95 % des gehörten Textes erreicht. Nach weiterer Gewöhnungszeit von 3 bis 4 Tagen bei täglichem fünfzehnminütigem Hören steigt die Verständlichkeit auf ca. 99 %. Der Gewöhnungsvorgang gleicht dem Einhören in die Sprechweise einer Person, deren Sprache stark dialektgefärbt ist."

Bei realen Vorleseautomaten ergeben sich also für geübte und ungeübte Zuhörer unterschiedliche Verständlichkeiten. Dies hat zwar den Vorteil, daß der Einsatz auch von noch nicht idealen Vorleseautomaten in Anwendungen mit kooperationsbereiten Zuhörern durchaus möglich ist. Entwicklungsziel kann aber nur ein idealer Vorleseautomat sein, dessen Verständlichkeit von Anfang an gleich der Verständlichkeit von ungestörter menschlicher Sprache ist.

Durch diese Forderung ist zunächst die Art der Sprache festgelegt, denn um im gesamten deutschsprachigen Raum gleich gut verstanden zu werden, muß deutsche Hochlautung erzeugt werden, so wie sie z.B. von einem geschulten Sprecher gesprochen wird.

Die Regeln zur Erzeugung von Lautschrift und Akzent aus Text

sind zwar sehr komplex, aber weitgehend bekannt. Große Kenntnislücken gibt es dagegen noch bei der Umsetzung dieser Größen in eine Parameterform zur Ansteuerung von z.B. Formant- oder LPC-Synthetisatoren. Um Fehler zu vermeiden, behilft man sich meist damit, diesen Teil der Umwandlung anhand des Vorbilds eines bestimmten menschlichen Sprechers zu entwickeln (Allen, Hunnicutt und Klatt, 1979). Dabei wird auch die vom Vorbild erzeugte sprecherspezifische Information übernommen, die teils durch den spezifischen Aufbau von Glottis und Sprechtrakt, teils durch die sprecherspezifische Ausprägung der Phoneme und der intonatorischen Parameter gegeben ist, denn es ist noch nicht möglich, alle sprecherspezifischen von den sprachspezifischen Parametern zu trennen. Diese Vorgehensweise führt dazu, daß die Parameter, von denen Verständlichkeit und Natürlichkeit abhängen, nicht mehr getrennt kontrollierbar sind. Maßnahmen zur Erhöhung der Verständlichkeit beeinflussen gleichzeitig die Natürlichkeit und umgekehrt.

2.1.2 Natürlichkeit

Natürlichkeit ist eine weitaus komplexere Größe als Verständlichkeit. Bereits die Definition dieses Gütemerkmals fällt schwer. Während man z.B. bei Ansageautomaten Natürlichkeit als das Merkmal der synthetischen Sprache definieren kann, das die Originaltreue kennzeichnet, ist bei Vorleseautomaten Natürlichkeit nur als die Wahrscheinlichkeit definierbar, mit der Zuhörer die ausgegebene Sprache für die eines menschlichen Sprechers halten (Blauert, Kulas, Rühl und Schaffert, 1984).

Die im letzten Abschnitt beschriebene Orientierung von Vorleseautomaten an einem menschlichen Vorbild bietet deshalb den großen Vorteil, daß jederzeit "Originale" angefertigt werden können, mit denen man synthetisierte Kopien vergleichen kann. Ein Übertragen aller Eigenschaften des Vorbilds auf den Vorleseautomaten hat aber auch Nachteile. Wie eine Anwendungsstudie mit synthetischer Sprache zur Fahrplanauskunft (Deutsche Bundesbahn, 1980) gezeigt hat, wollen die Benutzer zwar eine gute Verständlichkeit, aber man soll an der Sprache noch erkennen können, daß sie nicht von

einem Menschen, sondern von einer Maschine erzeugt wurde. Eine zu natürliche Sprache wurde von den Benutzern als unheimlich empfunden und weniger akzeptiert als eine Sprache mit maschinellem Klang.

Eine wichtige Kenngröße der Natürlichkeit bei Systemen zur Übertragung und Wiedergabe von Sprache ist die Identifizierbarkeit des Sprechers. Überträgt man diese Kenngröße auf Vorleseautomaten, heißt das, daß Stimmcharakter und Sprechweise des Vorleseautomaten zeitlich konstant sein müssen. Selbst wenn der Vorleseautomat Stimme und Sprechweise ändern kann (z.B. um Nachrichten von verschiedenen Quellen mit unterschiedlichen Stimmen auszusprechen), darf dies nur kontrolliert und keineswegs zufällig geschehen.

Natürlichkeit ist also insoweit sinnvoll, als sie die Verständlichkeit und die Akzeptanz eines Vorleseautomaten erhöht. Wird darauf geachtet, daß der Vorleseautomat als Maschine erkennbar und identifizierbar bleibt, kann man ihn ganz oder in Teilen als Modell eines spezifischen menschlichen Sprechers auslegen. Dies ist besonders dann vorteilhaft, wenn noch nicht exakt bekannte Verhaltensweisen, z.B. Teile der intonatorischen Steuerung, realisiert werden müssen.

Wie in den folgenden Abschnitten gezeigt wird, lassen sich aus der Forderung nach Natürlichkeit im Hinblick auf die Entwicklung eines Konzepts für Vorleseautomaten konkrete technische Anforderungen herleiten.

2.1.3 Echtzeit-Betrieb

Die Forderung nach Echtzeit-Betrieb folgt aus der Natürlichkeit, weil zur Natürlichkeit auch ein kontinuierlicher Sprachfluß gehört. Zwar ist natürliche Sprache von Pausen unterbrochen, aber diese Pausen sind semantisch und syntaktisch bestimmt und haben definierte Längen.

Echtzeit-Betrieb kann allerdings für verschiedene Anwendungen

unterschiedlich definiert werden. Bei monolog-orientierten Systemen wie z.B. einer Vorlesemaschine für Blinde bedeutet Echtzeit, daß Sprache kontinuierlich ausgegeben werden muß, es steht also für die Umsetzung von Text in Sprache genauso viel Zeit zur Verfügung wie für das eigentliche Aussprechen. Das wiederum bedeutet eine erforderliche Umsetzungsgeschwindigkeit von über 200 Wörtern pro Minute für schnell gesprochene Sprache.

Vorleseautomaten, die in dialog-orientierten Systemen eingesetzt werden, müssen dagegen nicht nur kontinuierlich sprechen, sondern auch möglichst schnell antworten. Bei einfachen Angaben sollte die Antwortzeit unter einer Sekunde liegen, und "leicht erlernbares Wissen" sollte nach spätestens ein bis zwei Sekunden ausgegeben werden (Zimmermann, 1983). Da die Länge der Antwortzeit nicht nur vom Vorleseautomaten abhängt, sondern auch der steuernde Rechner einige Zeit zur Erzeugung einer Antwort benötigen wird, sollte die Umsetzungsverzögerung des Vorleseautomaten deutlich unter einer Sekunde liegen, in Anwendungen, die eine schnelle Reaktion erfordern, sogar unter 100 ms.

In 100 ms einen vollständigen Satz in Ansteuerparameter für einen Phonemsynthetisator umzuwandeln, ist mit heutigen technischen Mitteln bei ausreichender Sprachqualität nicht möglich, aber eine Zeitspanne von einer Sekunde sollte für nicht zu lange oder zu komplizierte Sätze erreichbar sein. Es sollten sich auch Algorithmen zur Satzbetonung finden lassen, die einen Start der Sprachausgabe erlauben, ehe der Satz vollständig analysiert ist. Die Umsetzungsgeschwindigkeit sollte in jedem Fall bei mindestens 10 Wörtern pro Sekunde liegen, das entspricht der dreifachen Sprechgeschwindigkeit von schnell gesprochener Sprache.

2.1.4 Ermüdungsfreiheit

Viele menschliche Sprecher üben durch ihre Sprechweise einen ermüdenden und einschläfernden Einfluß auf ihre Zuhörer aus, wenn sie einen längeren Vortrag halten. Das Nachlassen der Konzentration als Folge eines monotonen Vortrags tritt in noch viel stärkerem Maße bei Vorleseautomaten auf, die keine oder nur eine mo-

notone Intonation erzeugen. Obwohl es möglich ist, Vorleseautomaten mit hoher Verständlichkeit zu entwickeln, die keine intonatorische Information ausgeben, ist für ein ermüdungsfreies Arbeiten eine möglichst abwechslungsreiche Intonation unabdingbar.

2.1.5 Anpassung der Umsetzungsstruktur an den Sprechvorgang beim Menschen

Kenntnisse über das Regelsystem, das der Erzeugung von Sprache zugrunde liegt, gibt es in zweifacher Form. In der ersten, dem Menschen immanenten Form wird es dauernd benutzt, um Sprache zu erzeugen oder zu verstehen. In dieser Form ist das Regelwerk jedem Menschen zumindest für seine Muttersprache so weit bekannt, daß er (sofern er lesen und schreiben kann) jeden geschriebenen Text aussprechen und jeden gesprochenen Text in eine schriftliche Form umsetzen kann. Die andere Form ist die in der Wissenschaft übliche Beschreibung als Algorithmus. Da das Regelsystem zur Erzeugung von Sprache sehr komplex ist, wurde bisher noch kein Versuch unternommen, die Umsetzung durch einen umfassenden Algorithmus zu beschreiben, sondern sie wurde in Teilbereiche zergliedert, die mehr oder weniger vollständig beschrieben wurden.

Für einen Vorleseautomaten müssen also viele Algorithmen zu einem Regelwerk zusammengefaßt werden. Da die meisten dieser Algorithmen nur vorläufigen Charakter haben, ist es sehr wichtig, die Struktur des Gesamtsystems so zu halten, daß einzelne Teilsysteme leicht ausgetauscht oder geändert werden können. Die Gesamtstruktur des Systems, d.h. die Art und Weise, wie die Algorithmen zusammenarbeiten, sollte so weit wie möglich die Vorgänge im Menschen nachbilden, damit gesichert ist, daß Veränderungen in den Teilalgorithmen nicht zu unerwünschten Änderungen im Verhalten des gesamten Systems führen.

Für eine starke Orientierung der Umsetzung an den Vorgängen im Menschen gibt es noch einen weiteren Grund: die Aussprachefehler. Wenn die Strukturen der Umsetzung ähnlich den Vorgängen im Menschen gehalten werden, kann man annehmen, daß Umsetzungsfehler zu ähnlichen Ergebnissen wie menschliche Aussprachefehler führen.

Diese können dann vom Hörer einfacher korrigiert werden als Fehler, die von andersartigen Strukturen erzeugt werden.

2.1.6 Technische Einschränkungen

Neben den ergonomischen Anforderungen an einen Vorleseautomaten gibt es schon in der Einleitung erwähnte technische Einschränkungen. Ziel der Arbeit war es, ein System zu entwickeln, das möglichst sofort praktisch einsetzbar sein sollte, z.B. in der Blindenhilfe. Deshalb wurde nicht in einer höheren Programmiersprache und auf einem Groß- oder Prozessrechner ein "Laborsystem" entwickelt, sondern es wurde ein 8-Bit-Mikroprozessorsystem benutzt, basierend auf dem M6800 der Fa. Motorola. Um Probleme bei der Umsetzung in Echtzeit zu vermeiden, wurde als Programmiersprache Assembler verwendet. Da zu Beginn der Arbeit (1978) bereits Phonemsynthetisatoren kommerziell erhältlich waren, wurde die Entwicklung einer Umsetzung von Lautschrift in Ansteuerparameter für Synthetisatorfilter von vornherein nicht als Teil dieser Arbeit betrachtet.

2.2 Vorlesen beim Menschen

Der Vorgang beim Menschen, der der Umsetzung von Text in Sprache im Vorleseautomaten am ehesten vergleichbar ist, ist das Vorlesen eines Textes. Wäre der Vorleseautomat in der Lage, alle Vorgänge nachzuvollziehen, die sich dabei abspielen, würde als Endprodukt eine Sprache entstehen, die von der menschlichen nicht mehr unterscheidbar wäre. Die Psycholinguistik ist die Wissenschaft, die sich mit diesen Abläufen im Menschen bei der Erzeugung und beim Verstehen von Sprache befaßt. Leider hat sie bisher noch keine Modelle für die Spracherzeugung beim Vorlesen geliefert, sondern sie befaßt sich mit allgemeineren Vorgängen der Art

Idee	➔	Sprache	(Sprachproduktion)
Sprache	➔	Idee	(Spracherkennung)

Als Beispiel für diese Vorgänge soll hier der Ablauf der Sprachproduktion anhand des Modells von Clark und Clark (1977) vorgestellt werden, ähnliche Modelle finden sich bei Fromkin (1971, 1973) und Garrett (1975).

- Setze Kontext-Bezüge: Hier wird die Art und Weise festgelegt, wie gesprochen werden soll: ob es sich um einen Monolog oder einen Dialog, eine Frage, eine Antwort auf eine Frage, eine Aussage, eine Anweisung oder eine Beschreibung handeln soll. Die (noch nicht verbale) Botschaft wird also unter Berücksichtigung des Kontextes formuliert.
- Festlegen der Satzstruktur: Aufgrund der Kenntnisse des Sprechers darüber, was dem Zuhörer bekannt und was neu ist, werden Subjekt und Prädikat festgelegt und die Einzelheiten der Idee in Zusammenhang gebracht.
- Planung der Konstituenten: Wenn die Grundstruktur des Satzes festliegt, werden die Konstituenten oder Phrasen des Satzes in Wörtern ausformuliert. Konstituenten sind syntaktische Einheiten des Satzes wie z.B. Subjekt, Prädikat oder Objekt. Sie können aus nur einem einzigen Wort bestehen, aber es ist auch möglich, mehrere Einzel-Konstituenten zu einer einzigen zusammen zu fassen wie z.B. "der kleine dicke Mann da hinten an der Ecke".
- Festlegung des artikulatorischen Programms: Mit der Wahl von Wörtern werden diese zugleich in ein artikulatorisches Programm umgesetzt und in einem Zwischenspeicher abgelegt. Es enthält Phoneme und Betonungsmuster.
- Artikulation: Der letzte Schritt ist die Ausführung des artikulatorischen Programms. Dies wird von Mechanismen gemacht, die das Programm in eine zeitlich sequentielle Reihenfolge von Befehlen an die Artikulationsorgane umwandeln. Das Ergebnis ist die Sprache, die der Sprecher zu erzeugen beabsichtigte.

Dadurch, daß statt Text eine Idee als Eingangsgröße für das Modell verwendet wird, ergibt sich ein Ablauf, der nicht auf Vorleseautomaten übertragbar ist, weil die andere Eingangsvariable völlig andere Transformationen notwendig macht. Ähnlich ist es bei dem Ablauf der Erkennung von Text. Hierfür wurden in der Literatur keine Modelle gefunden. Da aber nicht anzunehmen ist, daß grundsätzlich andere Mechanismen wie bei der Erkennung gesprochener Sprache angewendet werden, läßt sich dafür ein analoges Modell ableiten:

- Ein einzelnes Wort wird gelesen und identifiziert. Bei der Spracherkennung muß hierzu in vielen Fällen der Kontext herangezogen werden, denn Sprache ist ein asynchrones Signal, und die Formatierung ist bei weitem komplexer als die in der Technik übliche Kennzeichnung durch feste Datenstrukturen, Start- und Stopbits. Auf der Textebene sind jedoch diese Formatierungen, nämlich die Wortzwischenräume, vorhanden, so daß der Mensch jedes einzelne Wort ohne Kontextkenntnis anhand seines mentalen Lexikons identifizieren kann.
- Konstituentenerkennung: Einzelne Wörter werden so lange gelesen, bis aufgrund des grammatikalischen Wissens des Lesers die Einzelwörter zu einer Konstituente zusammengefaßt werden können.
- Erkennung der Satzstruktur: Einzelne Konstituenten können entweder zu übergeordneten Konstituenten zusammengefaßt werden, oder sie werden so lange aneinandergereiht, bis sich aus der Summe der Konstituenten die Satzstruktur ergibt.
- Extrahieren der Idee: Aus der vollständig bekannten Satzstruktur und der Kenntnis der vorhergehenden Sätze kann die Idee, die der Erzeugung des Satzes zugrunde lag, entnommen werden.

Dieses Modell ist insofern vereinfacht, als Wechselwirkungen zwischen den einzelnen Verarbeitungsstufen vernachlässigt wurden;

so ist z.B. das Zusammenfassen von Einzelwörtern oder Konstituenten zu höher geordneten Einheiten vom Kontext und der zugrunde liegenden Idee abhängig. Die Rückkoppelungen sind bei der Texterkennung jedoch bei weitem nicht so wichtig wie für die Spracherkennung, weil eine fehlerfreie Wortidentifizierung gegeben ist. Deshalb können sie in diesem Prinzip-Modell vernachlässigt werden.

Auch dieses Modell ist, für sich allein betrachtet, für einen Vorleseautomaten nicht brauchbar. Stellt man es aber, wie im Bild 2.1 gezeigt, in geeigneter Weise neben das Modell der Sprach-

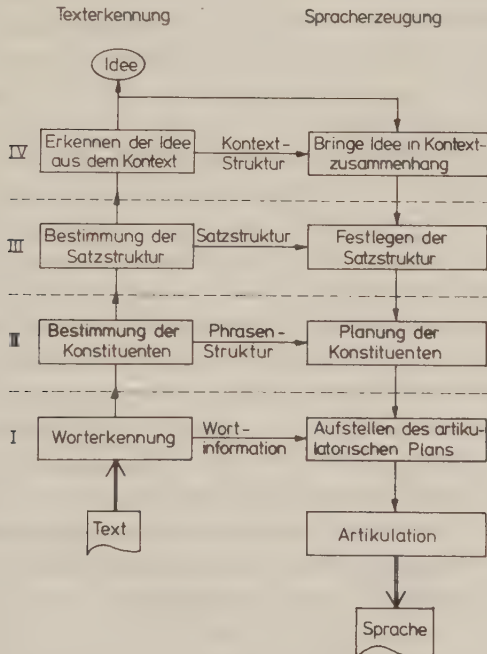


Bild 2.1: Zusammengefaßte Darstellung von Texterkennung und Spracherzeugung beim Menschen. Der erkennende Pfad entfernt bei der Abstrahierung auf jeder Ebene die Information aus der Nachricht, die der erzeugende Pfad zur Konkretisierung der Nachricht benötigt.

erzeugung, stellt man fest, daß sich beide Modelle sehr zweckmäßig ergänzen. Die Texterkennung wandelt den Satz in aufsteigender Richtung in eine immer abstraktere Form um. Als Nebenprodukt fällt dabei die grammatikalische Information an, die von den gegenüber liegenden Blöcken des Sprechmodells benötigt wird, um aus der Idee eine gesprochene Version des Satzes zu erzeugen. Die Kombination der beiden Modelle kann also als das Modell eines vorlesenden Menschen betrachtet werden.

2.3 Konzeptentwicklung für einen Vorleseautomaten

Durch die Zusammenfassung des erkennenden und des erzeugenden Modells zu einem Modell des Vorlesens vereinfacht sich der erzeugende Teil stark: Da der Erkennungsteil alle Information über die Struktur des Textes liefert, brauchen die Algorithmen des erzeugenden Zweiges diese Information nur noch in sprachrelevante Parameter umzusetzen. Alle Teile, die im erzeugenden Zweig zur Erzeugung der Struktur aus der Idee dienen, entfallen.

Trotz der Vereinfachung ist das Modell für Vorleseautomaten noch ungeeignet. Nach derzeitigem Stand der Technik ist es nämlich noch nicht uneingeschränkt möglich, bei unbeschränktem Wortschatz die "Idee" eines Satzes zu erkennen, also seinen Inhalt zu verstehen. Dies ist aber auch nicht notwendig, denn bereits ganz einfache Vorleseautomaten, die nur Information auf Wortebene auswerten, z.B. nur eine Lautschrift ohne Betonungsinformation nach Regeln erzeugen, können verständliche Sprache liefern. Die Idee ist auch dann schon in der phonematischen Beschreibung der Sprache enthalten, wenn der Vorleseautomat keine Information berücksichtigt, die auf höheren Ebenen erzeugt wird.

Daß die Kenntnis der Idee für die Aussprache von Sätzen nicht notwendig ist, darf nicht zu der Annahme verführen, daß die Idee die Aussprache des Satzes nicht beeinflusst. Ob emphatische Betonungen vorliegen, ob der Satz ironisch gemeint ist, ob bestimmte Zustände wie Trauer, Freude, Eile und ähnliches zum Ausdruck gebracht werden müssen, muß über den Inhalt des Satzes sowie den

Inhalt des Kontextes entschieden werden, die Form des Satzes gibt nur in Ausnahmefällen Hinweise. Da aber noch keine maschinellen Systeme existieren, die unbeschränkten Text verstehen können, ist es sinnvoll, sämtliche Verarbeitungsschritte auf den verschiedenen Ebenen, die einzig zum Gewinnen der Idee dienen, wegzulassen und nur Algorithmen zu berücksichtigen, die für die Aussprache wichtige Information erzeugen.

Durch den Verzicht auf die Bestimmung des Inhalts der Sätze wird der Vorleseautomat zwar erst realisierbar, die fehlende Information über Sprechweise und Empfindungen hat jedoch Rückwirkungen auf die Spracherzeugung, denn es kann nur eine "normierte" Sprechweise erzeugt werden. Dies kann z.B. eine neutrale Stimme mit mittlerer Sprechgeschwindigkeit sein. Auch wird es nur noch in Ausnahmefällen möglich sein, Kontextbezüge festzustellen, da deren Erkennung schon sehr stark mit der Erkennung der Idee verbunden ist.

Unter Berücksichtigung der Vereinfachungen läßt sich das Modell des menschlichen Vorlesens zu dem in Bild 2.2 gezeigten ersten Modell eines Vorleseautomaten vereinfachen. Dabei wurden auf allen Ebenen die analysierenden und die synthetisierenden Blöcke zusammengefaßt. Die Algorithmen der verschiedenen Ebenen legen ihre Ergebnisse in dem gemeinsamen Speicher namens 'artikulatorischer Plan' ab. Aus diesem Speicher wird die Sprache erzeugende Artikulation angesteuert, die als einziger selbständiger Teil des Syntheseteils übrig geblieben ist.

Da zur Erzeugung von Sprache kommerziell erhältliche Phonemsynthetisatoren verwendet werden sollen, vereinfacht sich die Artikulation, d.h. der Algorithmus zur Erzeugung der Artikulationsparameter erheblich, denn die Umwandlung der Lautschrift in Ansteuerparameter für den Syntheseteil eines Vocoders wird vom Synthetisator intern vorgenommen. Die Artikulation braucht nur noch die im artikulatorischen Plan enthaltene wortbezogene Lautschrift entsprechend der Satzbetonung zu korrigieren (d.h. Koartikulationen und Elisionen berücksichtigen) und Akzente sowie rhythmische Gliederung in Laut- und Pausendauern, Grundfrequenzverlauf und Lautintensität umzurechnen.

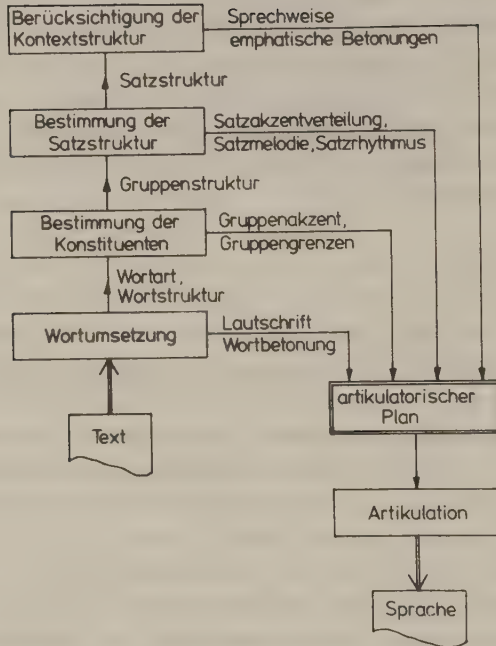


Bild 2.2: Vereinfachung des Modells des menschlichen Lesens und der Spracherzeugung als erstes Konzept eines Vorleseautomaten.

Ein großer Vorteil der in Bild 2.2 dargestellten Struktur ist, daß zur Erzeugung von Sprache nicht alle Ebenen voll ausgebaut sein müssen, sondern daß auch schon ein Minimalsystem Sprache liefern kann. Das heißt, selbst ein System, das nur eine Lautschrift völlig ohne Betonung erzeugt, kann eine zwar qualitativ eingeschränkte, aber verständliche Sprache liefern. Dadurch ist es möglich, das System unter Wahrung der Gesamtstruktur stufenweise auszubauen und trotzdem jederzeit ein lauffähiges System zu haben, was nicht nur zum Test neu hinzugefügter Algorithmen vorteilhaft ist.

Das Modell des Vorleseautomaten ist in der vorliegenden Form auf Text in Wortform abgestimmt. Um es universell einsetzbar zu

machen, muß der Wortumsetzung noch eine Textvorverarbeitung vorgeschaltet werden. Die Textvorverarbeitung hat die Aufgabe, Zahlen in verschiedenen Formaten, Sonderzeichen und Abkürzungen in eine "ausgeschriebene" Form zu bringen und die Sprachausgabe nach den Wünschen des Benutzers zu steuern.

Die Teilaufgaben der Vorverarbeitung lassen sich aus sprachwissenschaftlicher Sicht ohne weiteres in die Verarbeitungsebenen Wortumsetzung und Analyse der Satzstruktur einordnen. Vom ingenieurwissenschaftlichen Standpunkt ist es jedoch sinnvoll, die Textvorverarbeitung bereits auf der höchsten Konzeptebene als separaten Block darzustellen, denn ihre Aufgaben sind sehr stark

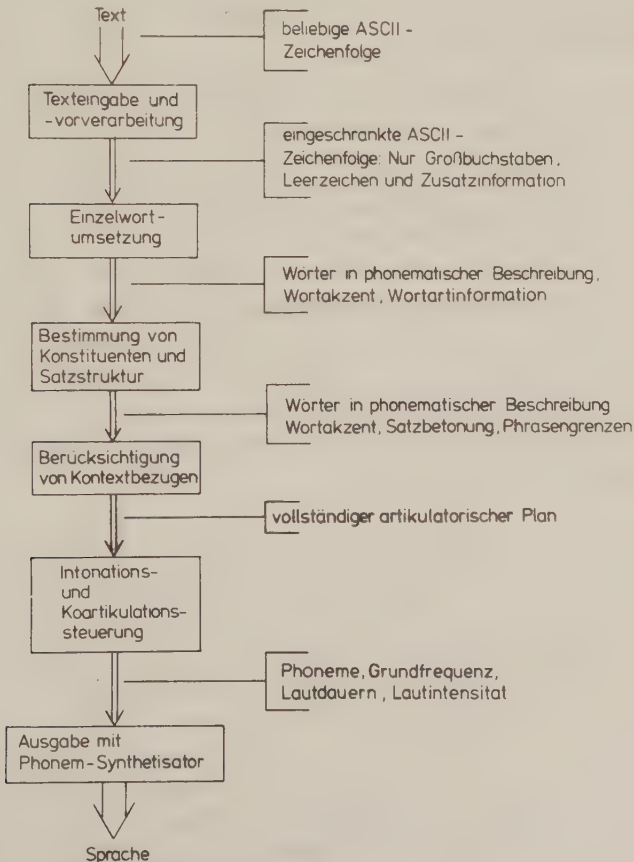


Bild 2.3: Grundkonzept des Vorleseautomaten

anwendungsbezogen. So sollte z.B. ein Vorleseautomat, der hauptsächlich Texte vorliest, 1983 als eine Jahreszahl, also als "neunzehnhundertdreiundachtzig" aussprechen. Wird der Vorleseautomat dagegen eingesetzt, um einem blinden Vermittler Eintragungen aus dem Telefonbuch vorzulesen, sollte die Zahl als "eins neun (Pause) acht drei" ausgesprochen werden.

Mit diesen Änderungen ergibt sich der in Bild 2.3 gezeigte Arbeitsablauf für den Vorleseautomaten.

2.4 Ein Vergleich mit einem anderen Modell für Vorlese- automaten

Am Massachusetts Institute of Technology (MIT) ist das MITalk-79 entwickelt worden, ein Vorleseautomat, der nach einem ähnlichen, in Bild 2.4 gezeigten Modell arbeitet. Von den eingezeichneten Blöcken I - IV der Umsetzung lassen sich drei, nämlich Textvorverarbeitung, Erzeugung der Intonation und der Synthetisatorteil, im vorgeschlagenen Grundkonzept wiederfinden. Aber Analyse auf Wort- und Satzebene, Transkription und Akzentbestimmung laufen anders ab, denn das Modell ist nicht am menschlichen Vorlesen orientiert, sondern ist als eine von praktischen Erwägungen diktierte Aneinanderreihung von Algorithmen der historisch gewachsenen Arbeitsgebiete Semantik, Syntaxlehre, Morphologie, Phonologie und Phonetik entstanden.

Wie das MITalk-79 bewiesen hat, muß eine Nichtbeachtung der vom menschlichen Verhalten her bekannten Schnittstellen nicht zu einer schlechten Sprachqualität führen. Um zu einer gut verständlichen Sprache zu kommen, ist es aber sicherer, die wenigen bekannten Schnittstellen zu beachten, als das Gesamtsystem wie eine "black box" zu behandeln und willkürlich zu zergliedern.

Dieses Vorgehen hindert nicht daran, feinere Gliederungen aus dem Modell des MITalk-79 in das vorgeschlagene Modell zu übernehmen, wie z.B. die Zerlegung der Transkription in Phonemisierung, morphembezogene und wortbezogene Koartikulationen. Sie sind je-

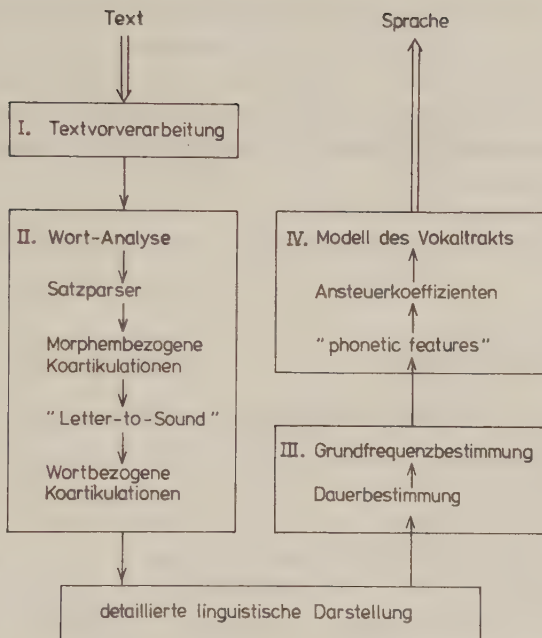


Bild 2.4: Ablauf der Umsetzung von Text in Sprache beim Vorleseautomaten MITalk-79 (nach Allen, 1981)

doch kein Teil des Grundkonzepts, sondern nur Arbeitshilfen zur weiteren Strukturierung der Algorithmen.

2.5 Einschränkung des Arbeitsumfangs

Wie weiter fortgeschrittene Projekte für amerikanisches Englisch (Allen, Hunnicutt und Klatt, 1979) und für Schwedisch (Carlson und Granström, 1976) zeigen, werden zur Realisierung von Konzepten vergleichbaren Umfangs mehr als 10 Mannjahre benötigt. Da dies den Rahmen einer Dissertation sprengt, beschränkt sich die vorliegende Arbeit im wesentlichen auf die Wortumsetzung, ergänzt durch die Textvorverarbeitung und die Verfahren, die zur Erzeugung einer Wortintonation benötigt werden. Auf Konstituen-

ten-, Satz- und Kontextebene wurden keine systematischen Arbeiten durchgeführt, sondern nur einige phänomenologische Maßnahmen getroffen, um die Sprachverständlichkeit der lauffähigen, aber unvollständigen Zwischenversionen des Vorleseautomaten zu erhöhen. Die noch ausstehenden Arbeiten werden im Rahmen einer Folgedissertation behandelt. Näheres darüber findet sich bei Kulas und Rühl (1982a, 1982b, 1984a, 1984b).

2.6 Zusammenfassung

Für einen Vorleseautomaten ergeben sich zwei Hauptforderungen:

- Verständlichkeit
- Natürlichkeit, soweit sie zur Verständlichkeit und Akzeptanz des Vorleseautomaten beiträgt.

Aus diesen Hauptforderungen lassen sich technische Anforderungen herleiten:

- Der Automat muß in Echtzeit arbeiten.
- Er darf die Zuhörer nicht ermüden.
- Seine Struktur sollte sich am Ablauf des menschlichen Sprechvorganges orientieren.

Ein Modell des menschlichen Sprechvorganges wurde vorgestellt, und aus bekannten Modellen des Sprachverstehens wurde ein Modell der menschlichen Texterkennung entwickelt. Die Modelle des Sprechvorganges und der Texterkennung wurden zu einem Modell des vorlesenden Menschen zusammengefaßt. Daraus wurde das in Bild 2.3 dargestellte Konzept eines Vorleseautomaten entwickelt.

3. DAS KONZEPT DER WORTUMSETZUNG

Wie aus dem Gesamtkonzept des Vorleseautomaten (Bild 2.3) hervorgeht, müssen auf Wortebene drei unterschiedliche Aufgaben gelöst werden:

- Die in Rechtschrift vorliegenden Wörter müssen in eine Lautschrift umgesetzt werden.
- Für die Wörter muß der Wortakzent bestimmt werden.
- Für (möglichst!) jedes Wort muß Information gewonnen werden, mit deren Hilfe auf den höheren Ebenen Phrasierung und Satzintonation bestimmt wird.

Die Weise, wie diese drei Aufgaben gelöst werden, hängt sehr stark von der Struktur der jeweiligen Sprache ab. Während z.B. die Rechtschreibung der finnischen, spanischen und italienischen Sprache sehr lautbezogen ist, läßt sich die deutsche Sprache nicht direkt von der orthographischen in eine phonematische Form umcodieren, denn die Ausspracheregeln, so wie sie z.B. im Duden (Mangold, 1973, 1974a, 1974b,) und bei Siebs (1969) angegeben sind, setzen einen "verständigen" Sprecher voraus, der Wörter in ihre Morpheme (Grundeinheiten wie Stämme; Vorsilben, Nachsilben und Endungen) zerlegen kann und damit die Struktur der Wörter erkennt. Deshalb sollen hier zunächst verschiedene Umsetzungsverfahren untersucht werden.

3.1 Prinzipielle Möglichkeiten der Wortumsetzung

Bei unbegrenztem Wortschatz bieten sich für die Umsetzung von orthographischem Text in eine phonematische Form nur zwei praktikable Transformationsebenen an: Wortebene und Morphemebene. Die auf diesen Ebenen aufbauenden Verfahren sind sich sehr ähnlich. Entweder werden möglichst komplette Listen aller möglichen Basiselemente einschließlich ihrer phonematischen Transkription aufge-

stellt, oder es wird über Regeln transformiert.

Eine Umsetzung auf Wortebene nur über Listen ist grundsätzlich fehlerfrei, aber nicht durchführbar, weil die Menge der deutschen Wörter offen ist. Pro Monat tauchen ca. 100-200 neue Wörter auf, so daß die Listen nie vollständig sein können. Vom Umfang her würde ein halbwegs vollständiges Lexikon unrealistisch groß, denn allein die Standard-Lexika Duden und Siebs haben mehr als 100 000 Eintragungen. Zusammen mit den Fachsprachen würde die Anzahl der Eintragungen in die Millionen gehen.

Eine Umsetzung auf Morphemebene über Listen ist nicht mehr grundsätzlich fehlerfrei, denn an den Morphemgrenzen auftretende Koartikulationen (man vergleiche die Aussprache des 'd' in: das Rad - des Rades) müssen nach Regeln korrigiert werden, außerdem muß auch die Tiefenstruktur des Wortes nach Regeln erkannt werden, um den Akzent richtig zu setzen. Die hierzu benötigten Listen sind kürzer, für deutschstämmige Wörter werden ca. 10000 Morpheme benötigt.

Damit sind aber nicht die fremdstämmigen Wörter abgedeckt, die zwar nur ca. 4% im laufenden Text ausmachen (Meyer, 1964), doch absolut Millionenhöhe an Zahl erreichen, und die durch ihre Herkunft aus unterschiedlichen Sprachen aus einer weit größeren Anzahl Morpheme als der deutschstämmige Wortschatz gebildet werden. Durch die unterschiedliche Herkunft sind die Kombinationsmöglichkeiten der Morpheme miteinander eingeschränkt, was durch Regeln erfaßt werden muß. Außerdem können neu auftauchende Wörter neue Morpheme enthalten.

Gegenüber den Umsetzungen nach Listen sind die Umsetzungen nach Regeln mit mehr Fehlern behaftet, aber dafür ist der Wortschatz unbeschränkt.

Eine Umsetzung nach Regeln auf Wortebene hat gegenüber der auf Morphemebene den Vorteil, daß die Wörter als umzusetzende Einheiten sehr einfach erkannt werden können, denn sie werden von Leerzeichen oder Interpunktionen begrenzt. Dafür ist die Fehlerquote aber hoch, denn Fehler an den Wortgrenzen von zusammenge-

	U m s e t z u n g			
	mit Listen		nach Regeln	
	auf der Ebene		auf der Ebene	
	Wort	Morphem	Wort	Morphem
prinzipiell un- beschr. Wortschatz	nein	nein	ja	ja
Bestimmung der Lautschrift über	Listen	Listen	Regeln	Regeln
Bestimmung des Akzents durch	Listen	Listen	Regeln	Regeln
Behandlung der Ko- artikulationen auf Wortebene durch	Listen	Regeln	Regeln	Regeln
auf Satzebene durch	Regeln	Regeln	Regeln	Regeln
Bestimmung der Wortart über	Listen	Listen+ Regeln	nicht möglich	z. Teil möglich
Fehlerquote	0	sehr kl.	groß	klein
Aufwand	sehr gr.	groß	klein	mäßig

Tab. 3.1: Vergleich der Eigenschaften verschiedener
Umsetzungsprinzipien

gesetzten Wörtern treten häufig auf. Soll die Fehlerquote verringert werden, steigt die Anzahl der Regeln sehr stark an. Eine Bestimmung des Akzents nach Regeln auf Wortebene ist nur schwer möglich, weil die Wortstruktur nicht bekannt ist.

Bei der Umsetzung nach Regeln auf Morphemebene ist die Erkennung der Morphemgrenzen schwierig. Die Umwandlung selbst ist jedoch einfach und hat eine geringe Fehlerquote, denn die Ausspracheregeln des Deutschen sind auf Morpheme bezogen. Im Ver-

gleich zur Umsetzung auf Wortebene liegt bei korrekt erkannter Wortstruktur und gleicher Anzahl Regeln die Fehlerquote bei der Erzeugung der Lautschrift weitaus niedriger, außerdem steigt der Aufwand für niedrigere Fehlerquoten weniger stark an.

Die Bestimmung des Wortakzents ist bei bekannter Struktur des Wortes nach Regeln durchführbar, und eine Bestimmung der Wortart ist teilweise möglich.

Die Eigenschaften der vier Umsetzungsprinzipien sind in Tab. 3.1 noch einmal gegenübergestellt.

In der Praxis werden fast nur Kombinationen von zwei oder mehr dieser vier Umsetzungsprinzipien angewendet, wobei jedoch meist eines dominiert und die weiteren nur zur Behandlung von Ausnahmen benötigt werden. Die Wahl des dominierenden Prinzips hängt von der Struktur der Sprache und der erwünschten Fehlerquote ab. Die für die deutsche Sprache realisierten Verfahren sowie einige Verfahren für andere Sprachen sollen jetzt vorgestellt werden.

3.2 Ein Überblick über realisierte Verfahren zur Wortumsetzung

3.2.1 Umsetzung mit Listen

Transkriptionen über Listen können prinzipiell keinen unbeschränkten Wortschatz umsetzen. Sie werden deshalb in Vorleseautomaten nur in Kombination mit Verfahren nach Regeln zur Behandlung der nicht in den Listen enthaltenen Ausnahmen eingesetzt.

Wegen des enorm hohen Aufwands wurde bisher noch kein Vorleseautomat entwickelt, der vornehmlich nach Listen auf der Wortebene umsetzt. Wortlisten werden aber insbesondere in der deutschen Sprache oft verwendet, um "häufige Ausnahmen" korrekt umzusetzen (Mangold und Stall, 1978; Kommenda, 1983; Breuer et. al.; 1979).

Auf Morphemebene wurde von J. Allen (Allen, 1973; Allen, Hunnicutt und Klatt, 1979; Allen, 1981) ein Algorithmus für amerikanisches Englisch entwickelt. Mit Hilfe einer Liste von ca. 10 000 Morphemen kann dieser Algorithmus die 100 000 häufigsten Wörter des Englischen mit einer sehr geringen Fehlerquote zerlegen. Die anschließende Umsetzung in Lautschrift und Akzent wird teils über Listen, teils über Regeln durchgeführt. Regeln setzen auch nicht erkannte Wortbestandteile um. Obwohl das Laborsystem zu sehr niedrigen Fehlerquoten führte, wurde in den praktischen Realisierungen (Klatt, 1982; Groner et. al., 1982) wegen des hohen Aufwands ganz oder teilweise auf die Morphemlisten verzichtet.

In den Verfahren für die deutsche Sprache (Mangold, Stall und Zelinski, 1978; Grossmann, 1981; Breuer et. al., 1979) werden Morphemlisten als Ausnahmeverfahren eingesetzt, um spezielle Morpheme, d.h. Vorsilben, Nachsilben und Endungen zu erkennen. Nur bei dem Verfahren von AEG-Telefunken werden auch einige Stämme über Listen erkannt, die ohne Erkennung fehlerhaft umgesetzt würden (Mangold und Stall, 1978).

3.2.2 Umsetzung nach Regeln auf Wortebene

Die Umsetzung nach Regeln auf Wortebene bietet sich besonders für Sprachen mit phonetischer Schrift wie Italienisch, Spanisch und Finnisch an, denn mit einem kleinen Regelsatz können sehr niedrige Fehlerquoten erzielt werden (Santos und Nombela, 1982; Delmonte, 1981a, 1981b, 1982; Synte, 1980). Für die italienische Sprache erreichte Delmonte (1981b) z.B. 72 Fehler in einem eine Million Wörter umfassenden fortlaufenden Text.

Aber auch in Sprachen mit einer nicht lautnahen Schrift werden Regelverfahren auf Wortebene häufig wegen ihrer Einfachheit angewendet. So wurden für amerikanisches Englisch mehrere Algorithmen (McIlroy, 1974; Elovitz et al., 1976) entwickelt, von denen z.B. der NRL-Algorithmus von Elovitz eine Umsetzung von Text in Lautschrift mit 329 Regeln durchführt und dabei eine Fehlerquote von 10% bezogen auf einen Corpus von 50 000 Wörtern erreicht.

In Deutschland hat Slaby (1978) an einem Verfahren gearbeitet, mit dem nur auf Wortebene; d.h. ohne Hilfe von Ausnahmelisten und von Wortzerlegungen; Text in Lautschrift umgewandelt werden sollte. Wie Slaby selbst zu diesem Verfahren sagt; "... ist hier ein enormer Umfang des Regelsystems zu erwarten" (Slaby, 1978). Nach einer persönlichen Mitteilung von Slaby sind die Arbeiten an dem Verfahren eingestellt worden. Der Abschlußbericht über das Projekt wurde nicht veröffentlicht.

Ein ähnlicher Ansatz ist in einer Vorstudie zu dieser Arbeit (Rühl, 1977; Uhle, 1978) zu finden, jedoch durch Listen für Ausnahmen ergänzt. Hier wurde eine Fehlerquote von ca. 5 %, bezogen auf Laute, erreicht. Die Umsetzung erfolgte in Echtzeit auf einem Mikroprozessorsystem mit 1 kByte Programmumfang.

Einen Schritt weiter in Richtung einer Zerlegung nach Regeln auf Morphembasis geht das von Berry-Rogghe entwickelte Programm PHONOL (Berry-Rogghe, 1976). Hier werden vor der Umsetzung des Wortes von Text in Lautschrift einige Vor- und Nachsilben vom Wort abgetrennt, die zu fehlerhaften Umsetzungen führen können. Das Programm wurde in Fortran geschrieben und erreicht eine Fehlerquote von 2 %. An der Kernforschungsanlage Jülich und dem Institut für Kommunikationsforschung und Phonetik in Bonn wurde dieses Programm unter dem Namen PHONTEXT weiter verbessert, sowie durch eine einfache Wortbetonung und Textvorverarbeitung ergänzt (Brustkern und Willee, 1979; Breuer et al., 1979). Es belegt im verbesserten Zustand 62 kByte bei einer Fehlerquote von ca. 10 % (*) und arbeitet in Echtzeit.

(*) Man beachte die PHONTEXT-Fehlerquote von 10 % gegenüber der PHONOL-Fehlerquote von 2 %. Ursache für diesen Widerspruch und für die Nicht-Vergleichbarkeit von Fehlerquoten im Allgemeinen ist die Tatsache, daß Autoren nur sehr selten angeben, ob sich ihre Angaben auf falsch phonemisierte Wörter oder auf falsch erzeugte Lautzeichen (Phoneme) beziehen. Außerdem kann die Zählgrundlage (z.B.: nur eine dem Duden entsprechende Aussprache zulässig; umgangssprachliche oder dialektale Abweichungen zulässig; alle Abweichungen zulässig, die das Wort nicht unverständlich machen) die Fehlerquote stark beeinflussen. Soweit bekannt, wird

Auf einer noch morphemnäheren Ebene arbeitet das Verfahren von Grossmann (Grossmann, 1979, 1981, 1982; Grossmann et al., 1980). Wörter werden zwar nicht vollständig zerlegt, aber es werden Vorsilben (V), Nachsilben (N) und Endungen (E) mit Hilfe von Listen abgespalten. Z.B. wird "Veränderungsmöglichkeiten" zerlegt in

"(V)VER-(R)ÄNDERUNGSMÖG-(N)LICH-(N)KEIT-(E)EN.

Der Wortrest (R) wird dann mit wortbezogenen Regeln in Lautschrift umgesetzt. Das Regelsystem umfaßt ca. 1000 Regeln, und die gesamte Transkription belegt ca. 8 kByte auf einem Mikroprozessorsystem. Sie wird ergänzt durch eine Ausnahmenerkennung für falsch transkribierte Wörter und ist auf die häufigsten 8000 deutschen und 4000 fremdstämmigen Wörter optimiert. Für diese Wörter erreicht sie Fehlerquoten von 1 % bzw. 3 %. Es wird nur eine rudimentäre Wortbetonung und keine Satzbetonung erzeugt. Das System arbeitet in Echtzeit.

3.2.3 Umsetzung nach Regeln auf Morphemebene

Unter dem Namen "rede" wurde bei der Gesellschaft für Mathematik und Datenverarbeitung in St. Augustin (GMD) auf der Grundlage von PHONTEXT ein Umsetzungs-Programmsystem entwickelt, das auch eine einfache Satzbetonung enthält (Müller, 1981). Zusätzlich zur Affix- und Endungserkennung ist hier auch ein Suchalgorithmus für bestimmte Grenzen zwischen Stämmen, eingebaut, um 'sp', 'st', 'dr', 'br', 'gr', 'bl', 'gl', 'bn' und 'gn' im Wort korrekt umsetzen zu können. Da dies nicht in Form von Umsetzungsregeln auf Wortebene geschieht, sondern durch einen der Umsetzung vorgeschalteten Parser, liegt das Umsetzungsverfahren näher an der Morphem- als an der Wortebene. Das sehr umfangreiche Programm wurde mit Hilfe des Parser-Generators LEX geschrieben und läuft in Echtzeit. Über die Fehlerquote werden keine Aussagen gemacht.

Nicht nur eine teilweise, sondern eine vollständige morphologische Analyse führt das System SPRAUS-VS (Sprachausgabe mit Voll-

deshalb bei den Fehlerquoten die Meßmethode angegeben.

synthese) durch, das von D. Stall entwickelt wurde (Mangold und Stall, 1978; Mangold, Stall und Zelinski, 1978; Workshop am Heinrich-Hertz-Institut, 1979; Class et al., 1981). Dazu wurden "Mikrogrammatiken" für deutschstämmige (Stall, 1974, 1977) und fremdstämmige (Stall, 1974) Wörter entwickelt, die dann mit verschiedenen Umsetzungen (Stall, 1976) in Lautschrift und Wortbetonung überführt werden. Ein Schiedsrichteralgorithmus entscheidet darüber, ob der Zerlegungsalgorithmus für deutschstämmige oder fremdstämmige Wörter angewendet werden muß. Die Mikrogrammatiken erzeugen nach Auskunft von D. Stall auch Information für die Satzbetonung.

Das Programm benutzt Lexika verschiedener Art mit insgesamt etwa 1500 Eintragungen. Es wurde zunächst in SNOBOL-4 realisiert, belegte ca. 480k Worte Kernspeicher einer TR440 und lief nicht in Echtzeit (ca. 3 s/Wort nach Mangold, Stall und Zelinski (1978)). Seit 1981 läuft es auch in Echtzeit auf einer PDP 11/50 und belegt ca. 60 kByte (Kommenda, 1983). Die Fehlerquote beträgt 5 % bezogen auf 7951 Testwörter der Sprachstatistik von Meyer (1964) bzw. 2 % bei Wichtung mit der Sprachhäufigkeit (Class et al., 1981).

3.3 Ein Vergleich der Umsetzungen auf Wortebene und auf Morphemebene

Zum Zeitpunkt der Entwicklung des Konzepts der vorliegenden Arbeit waren für die deutsche Sprache nur die Arbeiten von Kästner (1972), Berry-Rogghe (1976), Stall (1974; 1976, 1977), die Vorstudie am Lehrstuhl sowie aus persönlichem Gespräch der Ansatz von Grossmann bekannt. Alle Systeme führten nur eine Umsetzung von Text in Lautschrift ohne Bestimmung des Wortakzents durch. In diesen Ansätzen zeigte sich, daß die Systeme mit Umsetzung nach Regeln auf Wortebene vom Konzept und vom Umfang her einfacher sind, wenn "kleine" Vorleseautomaten entwickelt werden sollen, die nur Lautschrift und ggf. eine Wortbetonung erzeugen. Bei hohen Ansprüchen bezüglich der Fehlerquote und im Hinblick auf

die Bestimmung zusätzlicher Wortinformation; wie sie für die Phrasierung eines Satzes und zur Bestimmung der Satzintonation benötigt wird, ist ein Konzept vorteilhafter; das auf einer Analyse der Wortstruktur zur Bestimmung der Morpheme basiert. Die Vorteile einer solchen Umsetzung auf Morphemebene sind:

- Aus der Forderung, die Umsetzungen möglichst an die Vorgänge im Menschen beim Vorlesen anzupassen; um so u.a. Fehler besser "verständlich" zu machen; folgt, daß eine Umsetzung mit Wortstruktur-Analyse weniger schwerwiegende Fehler macht, da sie die Vorgänge beim Menschen nachvollzieht.
- Um Wortakzente setzen zu können; sollte die Wortstruktur bekannt sein. Wie spätere Arbeiten (Breuer et al., 1979; Grossmann, 1982) gezeigt haben; ist es nur schwer möglich, auf Wortebene Akzente zu bestimmen.
- Es ergeben sich einfachere Regeln für die Erzeugung der Lautschrift gegenüber der Umsetzung auf Wortebene.
- Eine Wortstruktur-Analyse kann aus der Wortstruktur, nämlich aus Nachsilben und Endungen, sofern sie keine Stämme erkennt, Angaben über die Wortart machen, die für die Erzeugung einer Satzbetonung benötigt werden.

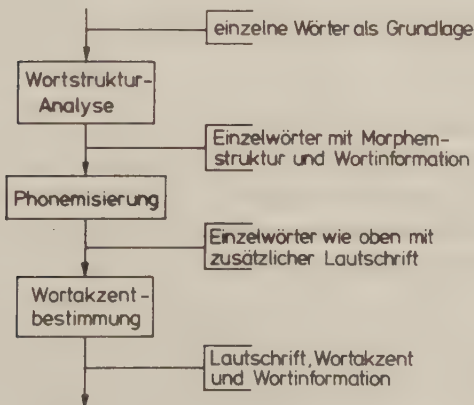


Bild 3.1: Erstes Konzept einer Wortumsetzung.

Wegen dieser Vorteile wurde in der vorliegenden Arbeit für den Vorleseautomaten ein Konzept entwickelt; das eine Analyse der Wortstruktur einschließt und die in Bild 3.1 gezeigte Grundstruktur hat.

3.4 Das Konzept der Wortstrukturanalyse

Eine Wortstrukturanalyse hat zwei Aufgaben: einerseits soll sie die Wörter in ihre Morpheme zerlegen und andererseits Information liefern, die von den Programmteilen zur Satz- und Phrasenbehandlung für die Bestimmung der Satzintonation benötigt wird.

Die zweite Aufgabe kann die Wortstrukturanalyse allerdings nur bedingt erfüllen. Anhand von Affixen und Endungen liefert sie zwar Information über die Sinnwörter (Substantive, Verben, Adjektive und Adverbien), ist aber nicht in der Lage, Information über die für die Satzbetonung weit wichtigeren Formwörter (Artikel, Präpositionen, Konjunktionen, Pronomen, Fragewörter und Hilfsverben) zu liefern, weil diese nicht nur klassifiziert, sondern identifiziert werden müssen. Deshalb muß für Formwörter eine Identifizierung über ein Lexikon vorgesehen werden. Eine lexikalische Identifizierung ist nur durchführbar, weil die Formwörter nicht den offenen Wortklassen angehören wie die Sinnwörter. Ihre Anzahl ist gering (ca. 150 - 600 je nach Zählweise) und dieses Wortinventar verändert sich nur sehr langsam im Lauf der Jahrhunderte.

Aus Gründen der Speicherökonomie ist es sinnvoll; nicht eine; sondern zwei Listen zu führen, da Pronomen und Hilfsverben flektiert bzw. konjugiert werden können. Für diese beiden Wortarten werden zunächst mögliche Endungen abgetrennt; erst der verbleibende Stamm wird anhand eines Stammlexikons identifiziert.

Beide Listen können nicht nur zur Identifizierung der Wortart, sondern auch zur Korrektur von fehlerhaft transkribierten Wörtern (sog. "Ausnahmen") verwendet werden. So werden Fehler bei häufig vorkommenden Wörtern vermieden.

Für die Hauptaufgabe der Wortstrukturanalyse, die Zerlegung von Wörtern in Morpheme, hat sich der von D. S. Stall entwickelte Ansatz zur Zerlegung von Wörtern mit Hilfe von Mikrogrammatiken im Hinblick auf eine möglichst fehlerfreie Dekomposition als geeignet erwiesen. Er ist aber sehr komplex und problematisch bei der Realisierung in Echtzeit. Deshalb wurde aufgrund eigener Erfahrungen mit dem Kästnerschen Algorithmus zur Phonemisierung von Wörtern (Rühl, 1977) auf der Grundlage der Arbeit von Kästner (1972) das Verfahren der Clusteranalyse entwickelt, das nicht wie der Ansatz von Stall am grammatikalischen Aufbau der Morpheme orientiert ist, sondern mit statistischen Regeln für Buchstabenfolgen Morphemgrenzen erkennt.

Die Funktion des Verfahrens wird im folgenden Kapitel beschrieben, hier sollen nur einige Eigenschaften erwähnt werden, die für das Konzept der Wortstrukturanalyse von Bedeutung sind. Die Clusteranalyse findet Morphemgrenzen in Wortzusammensetzungen. Sie ist nicht in der Lage, Morpheme zu klassifizieren, sie also als Stamm, Vorsilbe, Nachsilbe oder Endung zu erkennen, sondern kann nur Grenzen zwischen Morphemen feststellen. Da sie mit Regeln arbeitet, die auf einer Wortstatistik basieren, ist sie auch nicht in der Lage, alle Morphemgrenzen festzustellen. Grenzen können ziemlich sicher festgestellt werden, wenn sie zwischen zwei Stämmen liegen, dagegen werden öfter Grenzen zwischen Affixen oder Grenzen zwischen Affixen und Stämmen übersehen. Die Wahrscheinlichkeit, mindestens eine Morphemgrenze zu entdecken, wenn zwischen zwei Stämmen ein oder mehrere Affixe eingeschlossen sind, ist aber hoch.

Weil die Clusteranalyse nicht in der Lage ist, Morpheme zu klassifizieren, müssen Vorsilben, Nachsilben und Endungen durch Listenvergleich erkannt werden. Der Algorithmus zur Bestimmung von Affixen kann aber Ergebnisse der Clusteranalyse berücksichtigen, und die Suche nach Vor- oder Nachsilben braucht nicht nur an Wortanfang bzw. -ende anzusetzen, sondern kann auch die von der Clusteranalyse gefundenen potentiellen Morphemgrenzen als Ansatzpunkte verwenden. Aus diesem Grund wurde ein Ablauf der Strukturanalyse gewählt, wie er in Bild 3.2 zu sehen ist. Nachdem zunächst die Ganzwort-Ausnahmen erkannt worden sind, untersucht die

Klusteranalyse die restlichen Wörter und macht "Vorschläge", wo Morphemgrenzen liegen. Sämtliche weiteren Programmteile zur Bestimmung von Endungen, Vor- und Nachsilben beziehen diese Vorschläge der Klusteranalyse mit in ihre Berechnungen bei der Identifizierung und Verifizierung der entsprechenden Morpheme ein und korrigieren ggf. Grenzvorschläge, wenn sie zu widersprechenden Ergebnissen kommen.

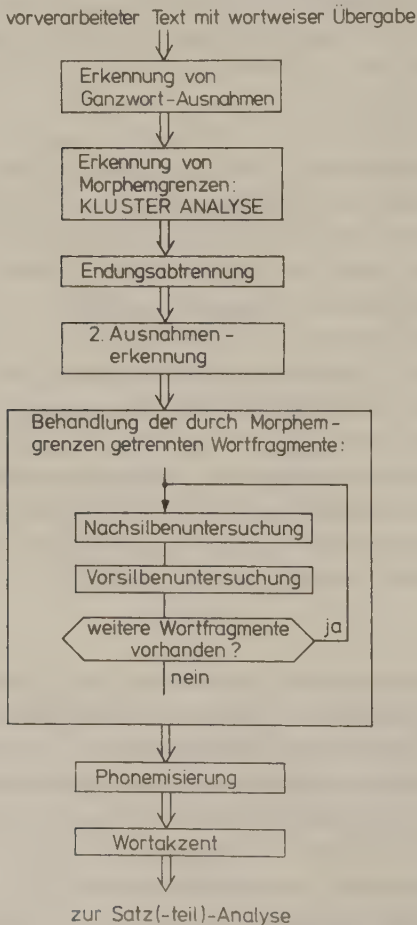


Bild 3.2: Aufbau der Wortumsetzung

Nach der Klusteranalyse werden zunächst die Endungen abgetrennt, die nur am Wortende, nicht aber innerhalb des Wortes bei Komposita auftreten können, und es wird untersucht, ob das Restwort im Stammllexikon eingetragen ist.

Wurde keine Ausnahme gefunden; wird versucht, Vor- und Nachsilben zu erkennen. Dabei wird zunächst von den sicher erkannten Grenzen ausgegangen, d.h. es wird am Wortanfang nach Vorsilben und vor evtl. erkannten Endungen nach Nachsilben gesucht. Erst wenn sich nach dieser Untersuchung noch von der Klusteranalyse vorgeschlagene Grenzen im Wortrest befinden, werden diese zum Ausgangspunkt für die Suche nach weiteren Affixen gemacht.

Hat ein Wort Nachsilben, kann oft über die Nachsilbe die Wortart bestimmt werden. Ebenso ist die Endungsabtrennung in der Lage, entsprechend der Endung Wahrscheinlichkeiten über Wortart und Kasus oder Konjugation anzugeben. Das ist eine Wortinformation, die für die Satzintonation über die offenen Wortklassen zur Verfügung stehen sollte.

3.5 Zusammenfassung

Mögliche Umsetzungsprinzipien und realisierte Verfahren zur Erzeugung von Lautschrift, Wortakzent und Wortinformation aus Text wurden verglichen. Dabei zeigte sich:

- Es ist nicht notwendig, als dominierendes Umsetzungsprinzip ein auf Listen basierendes Verfahren zu verwenden; denn die Rechtschreibung deutscher Wörter ist regelmäßig genug zur Anwendung von Regelverfahren.
- Verfahren zur Umsetzung nach Regeln auf Wortebene reichen aus, um "einfache" Systeme zu entwickeln; die nur Lautschrift und eine rudimentäre Wortbetonung erzeugen.
- Werden "höhere" Ansprüche an Lautschrift; Wortakzent und Wortinformation gestellt, ist es sinnvoll, eine Analyse der

Wortstruktur durchzuführen und nach Regeln auf der Morphem-ebene zu transkribieren.

- Da es nicht möglich ist, Wortinformation über die für die Satzintonation wichtigen Formwörter durch Regeln zu erhalten, müssen Formwörter über Listen erkannt werden.

Unter Berücksichtigung dieser Punkte wurde das in Bild 3.2 gezeigte Konzept der Wortumsetzung entwickelt. Im Mittelpunkt der Umsetzung steht dabei eine Bestimmung der Wortstruktur mittels Klusteranalyse zur Feststellung von Morphemgrenzen in zusammengesetzten Wörtern. Die Klusteranalyse arbeitet eng mit einer Endungserkennung und mit Verfahren zur Erkennung von Vor- und Nachsilben nach Listen zusammen. Erst durch die Zusammenarbeit von Affixerkennungen und der nach Regeln arbeitenden Klusteranalyse ist es möglich, mit großer Wahrscheinlichkeit den korrekten Wortaufbau festzustellen.

4. DIE ERKENNUNG VON MORPHEMGRENZEN

4.1 Ein linguistisches Modell deutscher Wörter

Im Rahmen der Strukturanalyse wurde von der folgenden allgemeinen Struktur für einfache Wörter ausgegangen:

$$W = \sum_{i=0}^n V_i + S + \sum_{k=0}^m N_k + \sum_{l=0}^p E_l \quad ; \quad n, m, p = 1, 2, 3, \dots \quad (I)$$

Ein Wort besteht demnach aus null oder mehr Präfixen V_i , einem Stamm, null oder mehr Affixen N_k und null oder mehr Endungen E_l . Einfache Wörter mit dieser Struktur können mit Hilfe von Vorsilben-, Nachsilben- und Endungserkennung vollständig zerlegt werden, der nach Abtrennung der Affixe nicht mehr weiter zerlegbare Wortrest muß notwendigerweise der Wortstamm sein.

Neben diesen einfachen Wörtern gibt es im Deutschen häufig Komposita, d.h. Wortzusammensetzungen mit mehreren Teilwörtern:

$$W_{\text{zus}} = \sum_{z=1}^r (W_z + F_z) + W_{r+1} \quad ; \quad r = 1, 2, 3, \dots \quad (II)$$

Die W_z sind Wörter mit Aufbau nach (I); die F_z sind Fugen, wobei auch die leere Fuge gültig ist. An dem Standardbeispiel für solche Zusammensetzungen, der

Donau-dampf-schiff-fahrt-s-gesellschaft-s-kapitän-s-mütze

sieht man, daß außer der leeren Fuge die Fuge 'S' sehr häufig auftritt. Ebenso ist am Beispiel zu sehen, daß Zusammensetzungen aus sehr vielen Teilwörtern bestehen können. Auch auf Zusammensetzungen läßt sich die für (I) angegebene Analyse mittels Affix- und Endungserkennungen anwenden, nur entspricht der nicht weiter zerlegbare Rest des Wortes nicht dem Stamm S; sondern einem Pseudostamm S' , der aus mehreren Morphemen zusammengesetzt ist.

Innerhalb des Pseudostamms S' treten zwischen zwei Stämmen S_z und S_{z+1} Terme der Form

$$\sum_{k_z=0}^{m_z} N_{k,z} + \sum_{l_z=0}^{p_z} E_{l,z} + F_z + \sum_{i_{z+1}=0}^{n_{z+1}} V_{i,z+1} \quad (III)$$

auf. Damit von einer Affixerkennung dieser Term aufgelöst werden kann, muß mindestens eine Morphemgrenze vor, in oder hinter dem Term erkannt werden.

Um solche Morphemgrenzen nach Regeln feststellen zu können, wurde im Rahmen dieser Arbeit ein Verfahren entwickelt, das auf Arbeiten von W. Kästner basiert.

4.2 Der Kästner'sche Ansatz zur Zerlegung von Wörtern

Der von Kästner (1972) vorgeschlagene Ansatz ist konzipiert als eine vollständige Zerlegung von deutschstämmigen Wörtern in Morpheme. Kästner selbst schreibt dazu:

"Wäre es möglich, ein vorgelegtes Wort so in Buchstaben-
gruppen zu segmentieren, daß man den einzelnen Gruppen Mor-
pheme zuordnen könnte; so kann man die im Duden sehr aus-
führlich dargelegten morphembezogenen Transformationsregeln
anwenden. Dieser Weg soll hier für die Transformation ge-
wählt werden. Die Transformation besteht daher aus zwei
Schritten:

- Ein vorgelegtes Wort wird durch Regeln so in Einheiten
zerlegt, daß man diesen Einheiten auf linguistischer
Ebene Morpheme zuordnen kann.
- Die im ersten Schritt gewonnenen Einheiten werden nach
Regeln transformiert."

Zur Zerlegung von Wörtern entwickelt Kästner ein Regelsystem, das sämtliche möglichen Morphemgrenzen innerhalb eines Wortes angibt. Grundlage dafür sind die Grapheme; d.h. die schriftlichen Codierungen der Laute der Sprache in der Rechtschrift. Dazu teilt er zunächst die Grapheme der deutschen Sprache in zwei Gruppen, nämlich in Vokalgrapheme (Vokale) und Konsonantengrapheme (Konsonanten) auf. Er bezeichnet dann in einem Wort jeweils die Gruppen von Vokalen bzw. Konsonanten, die direkt zusammenstehen; als Kluster und spezifiziert sie nach ihrer Stellung im Wort. In dem Wort 'Stauchen' kommen zum Beispiel vor:

'ST'	initiales Konsonantenkluster
'AU'	mediales Vokalkluster
'CH'	mediales Konsonantenkluster
'E'	mediales Vokalkluster
'N'	finale Konsonantenkluster

Grundkluster sind die Konsonanten- und Vokalkluster, die innerhalb eines Morphems auftreten können. Die Anzahl dieser Grundkluster ist beschränkt. Kästner fand insgesamt 13 initiale Vokalkluster, 16 mediale Vokalkluster; 13 finale Vokalkluster, 52 initiale Konsonantenkluster und 120 mediale bzw. finale Konsonantenkluster, die er leider nicht getrennt aufführt. Die verschiedenen Kluster sind in Anhang 2 aufgeführt.

Mit Hilfe der Grundklusterlisten können nun Wörter auf ihre Morphemstruktur untersucht werden.

Dazu bietet sich zunächst das von Kästner vorgeschlagene Verfahren an, über diese Listen sämtliche Stellen innerhalb eines Wortes zu bestimmen, an denen Morphemgrenzen liegen können. Morphemgrenzen können danach sämtliche Grenzen zwischen Klustern sein, außerdem können Morphemgrenzen in Medialklustern liegen, wenn das Medialkluster dadurch in ein Final- und ein Initialkluster zerfällt. Kennzeichnet man die potentiellen Morphemgrenzen durch einen Strich; ergibt sich z.B. bei dem Wort "Sonnenschein" mit der korrekten Zerlegung "Sonne-n-schein":

Son-n-e-n-sch-e-in

Man sieht, daß diese Art der Zerlegung ohne Zusatzregeln im allgemeinen keinen Schluß von einer möglichen Morphemgrenze auf eine reale Morphemgrenze erlaubt. Kästner gab deshalb seine ursprüngliche Absicht auf, eine vollständige Zerlegung von Wörtern anhand der Grundclusterlisten zu entwickeln; er beschränkte sich darauf, aus den Listen Regeln zu Umsetzung von Text in Lautschrift herzuleiten.

4.3 Die Clusteranalyse

4.3.1 Beschränkung der möglichen Morphemgrenzen

Da es nicht meine Absicht war, eine im linguistischen Sinn vollständige Wortanalyse zu entwickeln, sondern eine unter ingenieurmäßigen Gesichtspunkten zu möglichst wenig Aussprachefehlern führende, habe ich die Ergebnisse von Kästners Arbeit weiter untersucht. Um die Vielzahl der möglichen Morphemgrenzen, die Kästner angibt, einzuschränken, wurde zunächst mit Hilfe der Grundcluster untersucht, ob innerhalb von Wörtern Cluster sicher entdeckt werden können, die an eine Morphemgrenze anschließen oder in denen eine Morphemgrenze liegt. Ein solcher Fall liegt immer dann vor, wenn Cluster gefunden werden, die keine medialen Grundcluster sind. Diese Cluster werden im weiteren als Grenzbereiche bezeichnet. Die exakte Lage der Grenze in, vor oder nach dem Grenzbereich wird in diesem ersten Schritt noch nicht festgelegt.

Bei den Vokal-Grundclustern zeigte es sich; daß initiale und finale Grundcluster Untermengen der medialen Grundcluster sind. Das Vokalgrundcluster 'EI' z.B. kann sowohl am Morphemanfang, wie in "Eisen", in medialer Position wie in "Bein" als auch in finaler Stellung wie in "Brei" auftreten. Das bedeutet, daß anhand von Vokalzeichenfolgen, die innerhalb eines Wortes gefunden werden, keinerlei Aussage über Morphemgrenzen gemacht werden kann, auch wenn tatsächlich eine Morphemgrenze zwischen einem initialen oder finalen Vokalkcluster und einem Konsonantencluster liegen sollte wie z.B. in "Blech-eimer" oder "blei-haltig". Deshalb kann

anhand der medialen Vokalkluster nur dann ein Grenzbereich erkannt werden, wenn finale und initiale Vokalgrundkluster aneinander stoßen und beide zusammen kein mediales Vokalgrundkluster bilden wie z.B. in "Bau-er".

Bei den Konsonantenklustern trennt Kästner nicht zwischen medialen und finalen Grundklustern, sondern sagt nur, daß die medialen Kluster in der Menge der finalen Kluster enthalten sind. Aus diesem Grunde kann leider nicht mit der kürzeren Medialklusterliste gearbeitet werden, sondern es muß die 120 Eintragungen enthaltende Finalklusterliste verwendet werden. Diese konnte jedoch um 25 sehr seltene Ausnahmen gekürzt werden; die Kästner mit sämtlichen Vorkommen angegeben hatte, denn in diesen Beispielen traten die Kluster nur in finaler Position auf.

Die Mengen der finalen und initialen Konsonantenkluster unterscheiden sich wesentlich. Nur 17 einbuchstabige Kluster sowie die Kluster 'ST', 'SP' und 'SCH' sind in beiden Mengen enthalten. Außerdem folgt der Aufbau der initialen und finalen Kluster unterschiedlichen phonologischen Regeln. Dadurch ergibt sich eine hohe Wahrscheinlichkeit, daß Final- und Initialkluster einen Grenzbereich bilden, wenn sie über Morphemgrenzen hinweg aneinanderstoßen. Dies gilt auch für den Fall, daß zwei einbuchstabige Konsonantenkluster zusammenstoßen. Es gibt nämlich nur 58 mediale Konsonantengrundkluster, die durch Kombinationen von finalen und initialen Klustern gebildet werden können; aber allein durch Kombination der 16 finalen und initialen einbuchstabigen Kluster mit größerer Häufigkeit lassen sich 288 zweibuchstabige Kluster bilden. Also führen nur 13 % der Kombinationen zu Fehlern. Des weiteren läßt sich immer dann ein Grenzbereich bestimmen; wenn ein initiales Konsonantenkluster, das kein mediales Grundkluster ist, an ein finales Vokalkluster stößt.

Diese Betrachtung zeigt, daß die Wahrscheinlichkeit der Erkennung einer Morphemgrenze sehr stark davon abhängig ist; zwischen welchen Klustern sie sich befindet. Liegt sie zwischen Konsonantenklustern; besteht eine ziemlich große Wahrscheinlichkeit, daß sie entdeckt wird. Liegt sie zwischen einem finalen Konsonanten- und einem initialen Vokalkluster, wird sie nicht entdeckt. Eine

Grenze zwischen einem finalen Vokalkluster und einem initialen Konsonantenkluster hat sehr dagegen gute Chancen; entdeckt zu werden, wenn das Konsonantenkluster aus mindestens zwei Zeichen besteht. Das gleiche gilt für eine Grenze zwischen Vokalklustern, wenn mindestens ein Diphthong oder ein Umlaut beteiligt ist.

4.3.2 Abschätzung der Zulässigkeit der Beschränkung

Nachdem feststeht, daß die Wahrscheinlichkeit der Erkennung einer Morphemgrenze davon abhängig ist, ob ein Grenzbereich gebildet wird, wurden statistische Untersuchungen über den Aufbau der Morpheme des Deutschen durchgeführt, um festzustellen, wie häufig Grenzbereiche gebildet werden. Ausgehend von der o.a. Formel (III) in einer um die angrenzenden Stämme erweiterten Form

$$S_z + \sum_{k_z=0}^{m_z} N_{k,z} + \sum_{l_z=0}^{p_z} E_{l,z} + F_z + \sum_{i_{z+1}=0}^{n_{z+1}} V_{i,z+1} + S_{z+1} \quad (\text{IIIa})$$

lassen sich unter Vernachlässigung der möglichen Fuge $F_z = 'S'$ und unter Zusammenfassung von Nachsilben und Endungen unter dem Morphemtyp N die folgenden minimalen Morphemketten entwickeln:

$$\begin{array}{l} S_1 + S_2 \\ S_1 + N_1 + S_2 \quad S_1 + V_1 + S_2 \\ S_1 + N_1 + N_2 + S_2; \quad S_1 + N_1 + V_1 + S_2 \quad ; \quad S_1 + V_1 + V_2 + S_2 \end{array} \quad (\text{IV})$$

Innerhalb dieser und aller weiteren bildbaren Ketten können nun die folgenden Morphemtypen aneinandergrenzen:

$$\begin{array}{l} S + S \quad , \quad S + N \quad , \quad S + V \\ N + S \quad ; \quad N + N \quad ; \quad N + V \\ V + S \quad , \quad , \quad V + V \end{array} \quad (V)$$

Um für diese Morphemfolgen Aussagen über Grenzbereiche machen zu können, wurde auf statistische Ergebnisse über den Aufbau von deutschen Morphemen zurückgegriffen.

Erben (1980) untersucht einen Korpus von 37572 ein- und zweisilbigen Stämmen. Von diesen Stämmen fangen 4,2% mit einem Vokalkluster und 95,8% mit einem Konsonantenkluster an; am Stammende stehen zu 2,9% Vokale und zu 97,1% Konsonanten. Die entsprechenden Zahlen für Vorsilben und Nachsilben habe ich selbst anhand der im Umsetzungsprogramm verwendeten Affixlisten bestimmt. Damit ergibt sich:

	VV	VH	KV	KH
Stämme	4,2%	2,9%	95,8%	97,1%
Vorsilben	46,8%	21,8%	53,2%	78,2%
Nachsilben	44,8%	15,5%	55,2%	84,5%

dabei bedeuten:

VV : mit Vokal beginnend

VH : mit Vokal endend

KV : mit Konsonant beginnend

KH : mit Konsonant endend

Unter der grob vereinfachenden, aber für eine Abschätzung nützlichen Annahme einer Gleichverteilung der Morpheme innerhalb ihrer Klassen ergab sich für die Morphemketten nach (V) die in Tab. 4.1 gegebene Statistik des Buchstabenumfelds der Morphemgrenzen. Es zeigte sich, daß bei allen Ketten die meisten Grenzen zwischen Konsonantenklustern fallen, grenzen Stämme aneinander; ist dies sogar in über 90% aller Fälle so. Wenn rechts von der Grenze ein Stamm steht, ist die nicht erkennbare Grenze zwischen Konsonanten- und Vokalkluster kaum vertreten. Steht dagegen rechts der Grenze ein Affix; tritt sie in ca. 40% aller Fälle auf. Die nur in manchen Fällen erkennbaren Grenzen zwischen Vokal- und Konsonantenklustern und zwischen zwei Vokalklustern treten generell nicht sehr häufig auf.

Insgesamt gesehen; sind Grenzbereiche mit einiger Sicherheit zu finden, wenn rechts der Grenze ein Stamm liegt, Grenzen zwischen

	k + k	k + v	v + k	v + v	
S + S	93,0%	4,1%	2,7%	0,1%	S : Stamm
S + N	53,6%	43,3%	1,6%	1,3%	N : Nachsilbe
S + V	51,5%	45,7%	1,5%	1,3%	V : Vorsilbe
N + S	80,9%	3,5%	14,8%	0,7%	+ : Morphemgrenze
N + N	46,6%	37,9%	8,6%	6,9%	
N + V	44,7%	39,7%	8,2%	7,2%	k : Konsonanten- kluster
V + S	74,7%	3,3%	21,1%	0,9%	v : Vokalkluster
V + V	41,3%	36,7%	11,7%	10,3%	

Tab. 4.1: Häufigkeit des Auftretens der Morphemgrenze zwischen Konsonanten- und Vokalklustern in Abhängigkeit von den angrenzenden Morphemtypen.

Stämmen und Affixen rechts und Grenzen zwischen Affixen haben dagegen eine geringere Erkennungswahrscheinlichkeit.

Da Vorsilben und Nachsilben auch mit Hilfe von Listen erkannt werden können, ist es nicht nötig, alle Morphemgrenzen in (III) nach Regeln zu bestimmen; vielmehr genügt es, eine einzige zu erkennen. Ausgehend von dieser Grenze können dann nach links und nach rechts suchende Algorithmen mit den Listen alle Affixe bestimmen und so Morphemketten nach (III) analysieren.

Unter dieser Voraussetzung wurde abgeschätzt, wie groß die Wahrscheinlichkeit ist, eine Grenze in (III) exakt zu bestimmen. Dazu wurde zuerst die Annahme getroffen; daß die Lage der Morphemgrenze in Grenzbereichen vom Typ K+K mit 90% Wahrscheinlichkeit und in Grenzbereichen der Typen V+K und V+V mit 50% Wahrscheinlichkeit exakt bestimmt werden kann. Mit diesen Werten ergeben sich aus Tab. 4.1 Erkennungswahrscheinlichkeiten für Morphemgrenzen zwischen Morphemen des Typs:

S + S	:	85%
S + N	:	50%
S + V	:	48%
N + S	:	81%
N + N	:	50%
N + V	:	48%
V + S	:	78%
V + V	:	48%

Aus der Erkennungswahrscheinlichkeit für die Morphemgrenzen, an denen Affixe beteiligt sind, wird deutlich, daß eine zusätzliche Affixerkennung über Listen unbedingt notwendig ist, denn abgesehen von der nicht möglichen Klassifizierung der Affixe ist auch die Erkennung der Affixgrenzen weit weniger wahrscheinlich als die der Grenzen zwischen Stämmen.

Aus den o.a. Erkennungswahrscheinlichkeiten für einzelne Morphemgrenzen lassen sich Gesamtwahrscheinlichkeiten bestimmen, daß innerhalb einer Kette nach (IV) mindestens eine Grenze exakt erkannt wird.

S + S	:	85%
S + N + S	:	90%
S + V + S	:	89%
S + N + N + S	:	95%
S + N + V + S	:	94%
S + V + V + S	:	94%

Man sieht, daß mit steigender Komplexität des Terms (III) die Wahrscheinlichkeit, den Term korrekt aufzulösen, ansteigt. Für komplexere Realisierungen als die dargestellten läßt sich näherungsweise sagen, daß jedes weitere im Term eingefügte Affix die Wahrscheinlichkeit einer nicht korrekten Zerlegung halbiert.

Bei der Abschätzung wurde das Auftreten eines Fugen-'S' nicht berücksichtigt. Tritt eine solche Fuge auf, kann die Wahrscheinlichkeit, daß der Grenzbereich erkannt wird; nur vergrößert werden, denn außer bei den wenigen Morphemgrenzen zwischen Vokalklu-

stern verlängert und verändert die Fuge das oder die angrenzenden Konsonantenkluster. Je länger aber ein Konsonantenkluster ist, desto geringer ist die Wahrscheinlichkeit, daß es sich um ein mediales Grundkluster handelt. Da Fugen in der deutschen Sprache häufig sind, war zu erwarten, daß mehr Grenzbereiche erkannt werden als die obige Schätzung angibt.

4.3.3 Bestimmung der Morphemgrenze im Grenzbereich

Bisher wurde gezeigt, daß es möglich ist, allein durch Vergleich der medial auftretenden Kluster mit den medialen Grundklustern Grenzbereiche zu finden, in denen Morphemgrenzen liegen müssen. Für diese Bereiche muß in einem zweiten Schritt die genaue Lage der Morphemgrenze bestimmt werden. Zur Bestimmung der Grenzen vor oder in Konsonantenklustern wurden die Häufigkeiten benutzt, die Kästner zu jedem Konsonantenkluster angibt. Kästner verwendet eine Kennzeichnung der Häufigkeit mit 5 verschiedenen Klassen:

SS	=	sehr selten	(1-5)
S	=	selten	(6-60)
KV	=	kommt vor	(60-170)
H	=	häufig	(170-500)
SH	=	sehr häufig	(> 500)

Die Zahlen in Klammern geben die ungefähre Anzahl der Eintragungen für ein einzelnes Kluster in der jeweiligen Klasse an. Neben diesen Angaben kommen auch noch Zwischenwerte wie KV-H oder H-SH vor. Um diese Klassen zu quantifizieren, wurden den Klassen die folgenden Zahlen zugeordnet.

SS	:	1
SS-S	:	2
S	:	3
S-KV	:	4
KV	:	5
KV-H	:	6
H	:	7

H-SH : 8
SH : 9

Diese Zahlen entsprechen näherungsweise einer logarithmierten Häufigkeit und geben zugleich Wahrscheinlichkeiten für das Auftreten der Kluster an.

Wird nun ein mediales Konsonantenkluster entdeckt, das kein Grundkluster ist, kann es sich um ein Initialkluster handeln, eine Zusammensetzung von Final- und Initialkluster oder um eine Kombination mit einer Fuge zwischen den Klustern. Unter Berücksichtigung dieser drei Möglichkeiten werden alle möglichen Zerlegungen gebildet. Für jede mögliche Zerlegung werden die Wahrscheinlichkeiten für Initial- und Finalkluster addiert. Die so erhaltene Zahl ist ein Maß für die Gesamtwahrscheinlichkeit, daß diese Zerlegung richtig ist. Die Zerlegung mit der höchsten Zahl wird als wahrscheinlichste gewählt. Es folgen einige Beispiele für Analysen, wobei die Wahrscheinlichkeit einer Grenze unter dem ersten Zeichen nach der Grenze notiert wurde:

1. W U N S C H T R A U M
 -N -S -C -H -T -R

 0 0 0 0 9 0

2. W E I N F L A S C H E
 -N -F -L
 . . .
 . . .
 0 16 12

3. N A C H T R U H E
 N A C H T R A G
 -C -H -T -R

 0 0 16 16

Nicht mögliche Grenzen wurden mit '0' gekennzeichnet. Da bei einer vollständigen Zuordnung des Medialklusters zu einem Initialcluster keine Wahrscheinlichkeit für ein finales Konsonantencluster vorliegt; wird in diesem Fall die Wahrscheinlichkeit 'SS' = '9', also sehr häufig; für das finale Vokalkcluster berücksichtigt. Zum Beispiel:

4. B E G L E I T E N
 -G -L
 . .
 18 18

Wenn, wie in Beispiel 3. und 4.; zwei Zerlegungen gleiche Wahrscheinlichkeit haben, wird die Zerlegung mit der am weitesten links stehenden Grenze gewählt.

Da Fugen äußerst häufig sind; werden im Konsonantencluster auftretende 'S' immer als Fuge betrachtet, wenn sich dadurch eine gültige Zerlegung ergibt. Dies führt bei initialem 'ST' und 'SP' zu Aussprache Fehlern, z.B. wird 'FERNSTRASSE' in 'FERN-S-TRASSE' zerlegt. Die dadurch hervorgerufenen Aussprachefehler sind zwar störend, beeinträchtigen aber nicht die Verständlichkeit; weil sie mit einer dialektalen Variante (norddeutsch) zusammenfallen. **Diese Fehler** müssen in Kauf genommen werden, weil der umgekehrte Fall, d.h. die Zuordnung des Fugen-'S' zu einem mit 'T' beginnenden Initialcluster, wie z.B. in 'VOLKSTANZ', zu schwerwiegenden Aussprache Fehlern führen würde.

4.4 Fehler bei falschen Zerlegungen

Wie aus den Wahrscheinlichkeitsuntersuchungen über die korrekte Bestimmung aller Morphemgrenzen innerhalb des Terms (III) hervorgeht, bleibt ein Teil der Morphemgrenzen unerkannt; weil er nicht in einen Grenzbereich fällt; und ein Teil der Grenzbereiche wird falsch zerlegt. Da die Clusteranalyse auf die Zerlegung deutschstämmiger Wörter ausgelegt ist; ist außerdem zu erwarten; daß fremdstämmige Wörter falsch zerlegt werden, denn in Fremdstämmen

können Medialkluster auftreten, die nicht in der Liste der medialen Grundkluster enthalten sind. Die Clusteranalyse markiert dann Morphemgrenzen, die tatsächlich nicht vorhanden sind.

Für die Untersuchung der auftretenden Fehler wird nicht allein das Ergebnis der Wortanalyse herangezogen, sondern zusätzlich wird geprüft, wie sich die Fehler in der Aussprache bemerkbar machen, denn für die Verständlichkeit sind nicht die Fehler der Wortanalyse relevant, sondern erst die durch Fehler der Wortanalyse hervorgerufenen Aussprachefehler.

Zur Fehleruntersuchung wurden verschiedene Texte herangezogen, und zwar wurden fünf Artikel aus dem Lokalteil verschiedener Tageszeitungen, vier kurze Artikel aus der Frankfurter Allgemeinen Zeitung, ein längerer Artikel aus dem Spiegel; ein umfangreicher sozialwissenschaftlicher Bericht sowie Veröffentlichungen über Sprachsynthese verwendet. Ca. drei Viertel des Textumfangs hatte gehobenes bis wissenschaftliches Niveau; war also stark von Fremdwörtern und Zusammensetzungen durchsetzt. Für die Texte wurde zunächst eine Wortzählung durchgeführt; dabei wurden 10 647 Wörter gezählt. Von den 10 647 Wörtern kamen 2625 nur einmal vor, 1114 Wörter kamen zwei- oder mehrmals vor. Im Folgenden bedeutet eine Zahlenangabe XXX/YYY, daß XXX unterschiedliche Wörter zusammen YYY mal im Text auftraten.

Die Texte enthielten insgesamt 972/1394 Zusammensetzungen; davon wurden 633/903 oder 65 % gefunden und von der Clusteranalyse korrekt zerlegt, 235/360 oder 25 % wurden nicht gefunden und 104/131 oder 10 % wurden zwar gefunden; aber falsch zerlegt.

Gleiche Prozentzahlen ergeben sich, wenn man nur die zwei- oder mehrmals auftretenden Zusammensetzungen berücksichtigt. Für diese 193/612 mehrfach auftretenden Zusammensetzungen sind in Anhang 3 die 55/220 fehlerhaften Zerlegungen aufgeführt. Im Einzelnen traten folgende Fehler auf.

Der Wortakzent ist in den meisten Fällen von der nicht korrekten Zerlegung beeinflusst, aber nur in 4/11 Wörtern; deren erstes Teilwort fremdstämmig ist wie z.B. in 'Zivildienst', wird eine zu

betonende Endsilbe im ersten Teilwort nicht erkannt. Dies führt zu einem falschen Akzent und in den meisten Fällen zur Unverständlichkeit des Wortes. In den anderen Fällen führt die falsche Zerlegung nur zu einer weniger ausgeprägten Betonung, d.h. zu einem fehlenden Nebenakzent auf dem zweiten Teilwort wie z.B. in 'Fragebogen'. Diese Vereinfachung des Akzentmusters wird vom Hörer meist nur bemerkt; wenn er darauf hingewiesen wird. Die Verständlichkeit des Wortes ist zwar nicht so hoch wie die eines korrekt akzentuierten Wortes, aber höher als die eines monoton gesprochenen Wortes.

In 9/4 Fällen wird die Transkription von der fehlerhaften Zerlegung nicht berührt, die Wörter werden fehlerfrei ausgesprochen.

In 28/143 Fällen wird ein Schwa-Laut in einer Endung oder einem Präfix durch ein kurzes 'E' ersetzt, z.B. in 'Fragebogen'. Diese Art der Aussprache entspricht der sog. Überlautung; einer besonders deutlichen Aussprache zur Hebung der Verständlichkeit. Die Inkorrektheit mindert die Verständlichkeit nicht und fällt Hörern nur auf, wenn sie besonders darauf hingewiesen werden.

In 18/46 Wörtern wird vor dem mit einem Vokal beginnenden zweiten Teilwort kein Verschußlaut eingefügt; wie z.B. in 'Datenanalyse'. Da auch in der Umgangssprache dieser Knacklaut oft nur schwach oder gar nicht ausgeprägt ist; führt das Fehlen des Verschußlautes nur zu einer leichten Minderung der Verständlichkeit, der Fehler wird ohne Hinweis nur selten von Hörern bemerkt.

In 9/30 Wörtern wird 'ST' am Anfang des zweiten Teilwortes wie z.B. in 'vollständig' nicht wie 'SCH' sondern mit stimmlosem 'S' ausgesprochen. Diese Aussprache entspricht einer norddeutschen dialektgefärbten Aussprachevariante und ist in den meisten Fällen beabsichtigt (siehe weiter oben). Die Inkorrektheit der Aussprache ist auffällig, führt aber nicht zur Minderung der Verständlichkeit.

In 4/8 Wörtern wird ein Fugen-'S' zum Wortanfang des zweiten Wortes gezogen und als stimmhafter 'S'-Laut ausgesprochen, z.B. in 'Betreuung-saufgaben'. Der Fehler ist auffällig und macht

manchmal das Wort unverständlich.

In 4/11 Wörtern wird ein kurzer Vokal durch die falsche Morphemgrenze in einen langen Vokal überführt, z.B. in 'Fran-kreich'; ebenfalls ein auffälliger Fehler, der die Verständlichkeit des Wortes mindert.

Neben 8/19 Wörtern mit verminderter Verständlichkeit wurden 9/24 Wörter gefunden; deren Verständlichkeit stark verringert ist. Darin sind die 4/11 falsch akzentuierten Wörter enthalten. In 2/5 Fällen wurden statt stimmloser die entsprechenden stimmhaften Konsonanten erzeugt; in 2/6 Fällen wurde ein langer Vokal in einen kurzen überführt; und in 1/2 Fällen wurde statt eines kurzen 'E' ein 'Schwa' erzeugt.

Durch Fehler der Clusteranalyse an den 193/612 Wortgrenzen werden also nur 8/19 Wörter schwer verständlich und 9/24 Wörter unverständlich, das sind 4,1/3,1 % bzw. 4,6/3,9 %. Damit ergibt sich bei Zusammensetzungen eine Gesamtfehlerquote von 8,8/7,0 %.

Dieses auf den ersten Blick erstaunliche Mißverhältnis, daß durch ca. 35 % falsche Zerlegungen nur ca. 8 % Aussprachefehler hervorgerufen werden, liegt einerseits natürlich daran; daß die Affixerkennungen weitere von der Clusteranalyse nicht gefundene Morphemgrenzen erkennen; andererseits aber auch im Prinzip der Clusteranalyse begründet. Die Clusteranalyse findet sehr gut Morphemgrenzen; die in oder vor Konsonantenclustern mit mehreren Konsonanten liegen, dagegen keine Grenzen hinter Konsonantenclustern, keine vor oder nach einbuchstabigen Konsonantenclustern und nur einen Teil in zweibuchstabigen Konsonantenclustern.

Die Haupt-Fehlerart, die üblicherweise auftritt; wenn Wörter nach morphembezogenen Regeln auf Wortebene phonemisiert werden, ist die falsche Bestimmung der Vokallängen. Gerade dieser Fehler tritt im untersuchten Korpus nur 6/17 mal auf, weil dieser Fehler bei den Morphemgrenzen, die die Clusteranalyse nicht erkennt, nur an Grenzen in oder nach zweibuchstabigen Konsonantenclustern auftreten kann. Diese nicht erkannten Grenzen treten aber fast nur in Zusammenhang mit den Suffixen '-en' und '-er' auf, für deren

'E' keine kurz-lang-Entscheidung zu treffen ist, sondern eine die Verständlichkeit nicht beeinflussende Entscheidung zwischen kurzem 'E' und 'Schwa'.

Diese Endungen '-en' und '-er' sind es auch, die durch ihr überproportional häufiges Auftreten die Anzahl der nicht erkannten Grenzen gegenüber der Abschätzung in 4.3.2 in die Höhe getrieben haben. Es lassen sich zusätzliche Maßnahmen denken, um die Wahrscheinlichkeit der Erkennung von Grenzen nach den o.a. Endungen zu erhöhen, z.B. durch eine gesonderte Suche nach diesen Affixen mit anschließender Verifizierung der Grenzen, aber da an diesen Grenzen nur wenige schwere Aussprachefehler auftreten, ist dies beim derzeitigen Stand der Clusteranalyse nicht sinnvoll.

Sinnvoll sind aber Verbesserungen an einem zweiten Punkt, nämlich dort, wo Abschätzung und Meßergebnisse voneinander abweichen. In der Abschätzung war angenommen worden, daß in ca. 5 % aller Fälle die Lage der Morphemgrenze nicht exakt bestimmt werden kann, tatsächlich werden dagegen 10 % der Grenzen falsch lokalisiert. Davon sind 6,6 % Fälle, in denen ein 'S' falsch als Fugen-'S' bestimmt wird. Nimmt man in die Affixerkennung zusätzliche Regeln auf, die angeben, ob nach bestimmten Suffixen eine Fuge auftreten darf oder muß, würde dies für den vorliegenden Korpus die Quote der nicht korrekt bestimmten Morphemgrenzen um ein Drittel senken. Diese Maßnahme, die zugleich die Quote der Aussprachefehler um ein Viertel senkt, soll bei einer geplanten Vervollständigung des Datensatzes für die Affixerkennung berücksichtigt werden.

4.5 Fehler der Clusteranalyse in fremdstämmigen Wörtern

Die bisherigen Untersuchungen haben sich damit befaßt, die Fehler zu untersuchen, die an Grenzen in überhaupt nicht oder nicht korrekt erkannten deutschstämmigen, fremdstämmigen und gemischten Wortzusammensetzungen auftreten. Neben diesen Fehlern treten aber noch weitere Fehler dadurch auf, daß in Fremdwörtern mediale Grundcluster auftreten, die nicht in den deutschstämmigen media-

len Grundklustern enthalten sind.

Unter den 1114/8822 Wörtern des Korpus, die zwei oder mehrmals auftraten, sind 191/680 fremdstämmig. Da der korrekte morphologische Aufbau der Fremdwörter für die Transkriptionsprogramme nicht von Bedeutung ist, wurde nicht die korrekte Analyse der Fremdwörter überprüft, sondern es wurde nur untersucht, wie viele Fremdwörter durch Morphemgrenzen, die die Clusteranalyse in die Fremdwörter gelegt hat, falsch ausgesprochen werden.

Es wurden 14/42 falsch aufgesprochene Fremdwörter gezählt, d.h. bei 7,3/6,2 % der Fremdwörter führten Fehler der Clusteranalyse zu Aussprachefehlern. Fehler traten dabei hauptsächlich durch falsche Bestimmung der Vokallänge vor zweibuchstabigen medialen Konsonantenklustern auf. Typische fremdstämmige Grundcluster dieser Art sind z.B. 'th', 'bl', 'gr', 'mb', 'ph' und 'lv'.

4.6 Die Analyse von Vokalklustern

Wie Tabelle 4.1 zeigt, kommen Morphemgrenzen innerhalb von Vokalklustern zwar vor, sind aber selten. Um auch einen Teil dieser Grenzen erkennen zu können, wurden ein einfaches Verfahren entwickelt, das ähnlich wie die Clusteranalyse zunächst nach medialen Vokalklustern sucht, die keine Grundcluster sind. Werden solche Cluster gefunden, werden sie nach Regeln zerlegt.

Die Regeln führen aber keine Wahrscheinlichkeitsrechnung durch, sondern versuchen einfach, am Anfang oder am Ende desClusters einen Umlaut oder Diphtong zu identifizieren, durch den dann die Grenze festgelegt ist. Ist das nicht möglich, kann das Cluster nur aus zwei Zeichen bestehen, und die Grenze wird zwischen die zwei Zeichen gelegt.

Da als Vokalphoneme im An- oder Auslaut von deutschen Stämmen meist Diphtonge auftreten, funktioniert das Verfahren gut für deutschstämmige Wörter, die Morphemgrenze wird häufig korrekt bestimmt. Dagegen treten bei fremdstämmigen Wörtern mehr Fehler auf, denn es gibt weitaus mehr mediale Vokalgrundcluster als in

deutschstämmigen Wörtern; was zu falschen, weil nicht vorhandenen Morphemgrenzen führt. Quantitative Fehlermessungen sind nicht durchgeführt worden, denn aufgrund des sehr seltenen Auftretens von Grenzen in Vokalklustern war der verwendete Textkorpus zu klein, um aussagekräftig zu sein. Aus dem gleichen Grund ist eine Verbesserung des Verfahrens an dieser Stelle nicht vordringlich.

4.7 Zusammenfassung

Ausgehend von einem linguistischen Modell deutscher Wörter wurde mit Hilfe des Kästnerschen Verfahrens zur Umwandlung von Text in Lautschrift das Verfahren der Clusteranalyse entwickelt; mit dem Morphemgrenzen in Wortzusammensetzungen bestimmt werden können. Dieses Verfahren benötigt nur die Listen der initial und medial bzw. final auftretenden Grundkluster und arbeitet ansonsten sehr speicher- und zeitökonomisch nach Regeln.

Das Verfahren läuft in zwei Schritten ab: zuerst werden in Wortzusammensetzungen Grenzbereiche bestimmt, in denen Morphemgrenzen liegen müssen; dann wird innerhalb der Bereiche unter Berücksichtigung der Fugen die exakte Lage der Morphemgrenze bestimmt.

Da auf diese Weise nur die Bestimmung von Morphemgrenzen zwischen Wortstämmen hinreichend gut möglich ist, nicht aber die von Grenzen zwischen Affixen oder zwischen Affixen und Stämmen; muß die Clusteranalyse durch eine Affixerkennung unterstützt werden, die Vor- und Nachsilben in der Umgebung der gefundenen Morphemgrenzen identifiziert.

Weil die Clusteranalyse ein Algorithmus ist; dessen Regeln durch Statistiken des Morphemaufbaus ermittelt wurden, kann sie nicht fehlerfrei arbeiten. Es zeigt sich; daß das erstellte Verfahren nur zwei Drittel aller Grenzen korrekt bestimmt; 25 % aller Grenzen werden nicht gefunden. Dies führt jedoch nur zu 8 % verständlichkeitsmindernden Aussprachefehlern.

Damit ist die Klusteranalyse zwar als linguistischer Algorithmus zur vollständigen Analyse von deutschen Wörtern nicht geeignet, aber sehr gut als Instrument für die Sprachsynthese, denn wie eine Untersuchung der auftretenden Fehler gezeigt hat, findet sie Morphemgrenzen bevorzugt an den Stellen, wo morphembezogene Ausspracheregeln ohne Kenntnis der Morphemgrenze Fehler machen würden.

5. ERKENNUNG VON AFFIXEN UND AUSNAHMEN

Wie in den vorhergehenden Kapiteln beschrieben, kann die Klusteranalyse zwar viele Morphemgrenzen innerhalb von Worten finden, aber nicht alle. Insbesondere werden bei Affixen Morphemgrenzen übersehen oder falsch bestimmt. Außerdem ist die Klusteranalyse nicht in der Lage, die Morphemart zu bestimmen. Deshalb ist die Klusteranalyse allein als Instrument zur grammatikalischen Zerlegung von Wörtern nicht ausreichend.

Da nicht nur möglichst alle Morphemgrenzen erkannt werden sollen, sondern auch bestimmte Wortbestandteile in ihrer Morphemart als Vorsilbe, Nachsilbe oder Flexionsendung bestimmt werden müssen, ist es sinnvoll, diese Wortteile über Listen zu erkennen. Zusammen mit der Klusteranalyse bilden diese Erkennungen eine komplette Wortanalyse, die zwar nicht linguistischen Forderungen nach Vollständigkeit und Fehlerfreiheit genügt, aber dem praktischen Einsatz in einem auf Regeln basierenden Vorleseautomaten angemessen ist. Abgerundet wird das Verfahren zur grammatikalischen Analyse von Wörtern durch zwei Ausnahmen-Erkennungen.

5.1 Die Erkennung von Affixen

Affixe, d.h. Vorsilben und Nachsilben, müssen aus verschiedenen Gründen erkannt werden:

- Die Ausspracheregeln für Vorsilben sind nicht einheitlich und unterscheiden sich von denen für Stämme. Z.B. wird in 'ge-' und 'be-' das 'E' als Schwa ausgesprochen, in 'er-' und 'zer-' dagegen als kurzes 'E'.
- Die Grenzen vor und hinter Affixen müssen bekannt sein, weil sie die Aussprache der Affixe und der angrenzenden Morpheme beeinflussen. Beispiele: In 'Vorentscheidung' muß bekannt sein, daß zwischen 'vor' und 'ent' eine Morphemgrenze liegt, damit ein Knacklaut eingefügt wird, und in

'löslich' muß die Grenze zwischen 'S' und 'L' bekannt sein, damit das 'Ö' lang ausgesprochen wird.

- Es gibt im Deutschen drei verschiedene Arten von Affixen: Immer betonte, unter bestimmten Umständen betonte und nie betonte. Um den Wortakzent richtig bestimmen zu können, müssen Affixe deshalb nicht nur als Morphemart, sondern eindeutig als ein bestimmtes Affix identifiziert werden, damit die für dieses Affix gültigen Regeln zur Bestimmung des Akzents angewendet werden können.
- Aus den Nachsilben kann auf die für die Satzintonation wichtige Wortart geschlossen werden; z.B. sind auf 'heit', 'keit', 'nis' und 'ung' endende Wörter Substantive, dagegen auf 'bar', 'sam' und 'lich' endende Wörter Adjektive.

Zur Erkennung von Affixen werden zunächst Buchstabenfolgen am Wortanfang, Wortende oder an den von der Clusteranalyse vorgegebenen Stellen gesucht, die mit den Buchstabenfolgen in den Vorsilben- oder Nachsilbenlisten übereinstimmen. Dieser Identifizierung genannte Schritt ist im mathematischen Sinn notwendig, aber nicht hinreichend, denn viele mit Affixen identische Buchstabenfolgen können auch Teil eines Stammes sein. Beispiele:

<u>best</u> en	-	<u>best</u> immen
<u>Zungen</u>	-	Kreuz <u>ungen</u>

Deshalb ist ein Kontrollmechanismus notwendig; der untersucht, ob das als Buchstabenfolge gefundene potentielle Affix tatsächlich ein Affix ist. Dieser Verifizierung genannte Kontrollmechanismus prüft die Umgebung des potentiellen Affix nach folgenden Gesichtspunkten:

- Das zwischen dem potentiellen Affix und der nächsten Morphemgrenze liegende Restwort muß 'stammfähig' sein. Dazu wird untersucht; ob das Restwort aus mindestens einem Konsonantencluster und mindestens einem Vokalkcluster besteht, und ob es aus mindestens drei Buchstaben besteht. Durch die Beschränkung des Restwortes auf mindestens drei Buchstaben

werden in der Umgebung von zweibuchstabigen Stämmen Fehler gemacht. Zweibuchstabige Stämme treten im Deutschen aber so selten in Verbindung mit Affixen auf, daß diese Vereinfachung die Fehlerquote nur unwesentlich beeinflußt.

- Das Restwort nach einer Vorsilbe muß mit einem Initialkuster anfangen, bzw. das Restwort vor einer Nachsilbe muß mit einem Finalkuster enden. Diese Bedingung ist immer erfüllt, wenn die Grenze zwischen Affix und Restwort mit einer von der Clusteranalyse vorgeschlagenen Grenze zusammenfällt, sonst wird das an das Affix grenzende Cluster anhand der Grundkusterlisten überprüft.
- In zahlreichen Sonderfällen sind nur ganz bestimmte an das Affix angrenzende Buchstabenketten zugelassen bzw. andere ausgeschlossen.

Für die Durchführung von Identifizierung und Verifizierung bieten sich mehrere Möglichkeiten an. Im Hinblick auf Echtzeitbetrieb wäre es die schnellste und günstigste Lösung, wenn sowohl Identifizierung als auch Verifizierung von als Programm geschriebenen Algorithmen ausgeführt würden. Ein solches Vorgehen führt jedoch mit wachsender Affixzahl schnell zu unübersichtlichen, schlecht wartbaren und schwer erweiterbaren Programmen.

Vorteilhafter ist ein mit Tabellen arbeitendes Verfahren zur Identifizierung, denn Tabellen können sehr einfach korrigiert oder erweitert werden. Die Tabellensuche führt nicht zu Zeitproblemen, wenn ein genügend schnelles Suchverfahren benutzt wird. Deshalb wurde eine Identifizierung über Listen gewählt.

Verifizierung per Programm bedeutet, daß für jedes Affix eine Routine geschrieben werden muß, die die für das Affix spezifischen Untersuchungen im Umfeld durchführt. Eine Verifizierung per Programm ist ebenfalls unübersichtlich und schlecht wart- und veränderbar, wenn auch diese Nachteile durch die erzwungene programmtechnische Ausführung in vielen Einzelroutinen gemildert werden. Eine Verifizierung per Programm ist darum nur sinnvoll, wenn die Anzahl der nötigen Routinen sehr klein gehalten werden

kann. Untersuchungen zeigten, daß unterschiedliche Affixe nur selten gleiche Verifizierungen haben. Aber die Anzahl der notwendigen Routinen reduziert sich um ca. 40 % dadurch, daß für manche langen Affixe (vier oder mehr Buchstaben) eine Verifizierung ganz entfallen kann. Die Reduktion von 40 % war zu gering, als daß eine Verifizierung per Programm sinnvoll wäre. Deshalb wurde ein anderer Weg beschritten: Jedem Affix wird in der Liste eine Folge von Befehlen zugeordnet, die zur Verifikation interpretierend bearbeitet werden.

Wie die o.a. Aufzählung der bei einer Verifizierung zu erfüllenden Bedingungen zeigt, läßt sich für jedes Affix die Verifikation als eine Folge von Einzelbedingungen beschreiben, die erfüllt sein müssen, wenn tatsächlich ein Affix vorliegt. Unter den Bedingungen gibt es solche, die für alle Affixe erfüllt sein müssen wie z.B. die, daß das Restwort "stammfähig" sein muß, und es gibt affixspezifische Bedingungen wie die, daß in 'geben' die Zeichenfolge 'ge' kein Affix ist. Zur Beschreibung der affixspezifischen Bedingungen wurde ein Befehlssatz entwickelt, und es zeigte sich, daß alle Bedingungen mit einer kleinen Zahl von Befehlen beschrieben werden können.

Verifiziert wird nun, indem man sämtliche Bedingungen als Folge von Befehlen in den sog. Verifizierungseintrag schreibt, der jedem Affix zugeordnet ist. Dieser Eintrag wird nach der Identifizierung mit Hilfe eines Interpreters abgearbeitet, und er veranlaßt auch die immer durchzuführenden Überprüfungen. Als Ergebnis wird die Aussage geliefert, ob ein Affix vorliegt oder nicht.

5.1.1 Der Aufbau eines Verifizierungseintrags

Ehe auf die einzelnen Befehle eingegangen wird, soll zunächst der vollständige Aufbau eines Verifizierungseintrags erläutert werden. Der prinzipielle Aufbau eines solchen Eintrags ist für alle Affixe gleich und in Bild 5.1 abgebildet. Er besteht aus mindestens vier Feldern; deren erste drei und die Endekennzeichnung immer vorhanden sein müssen; die anderen sind optional.

In den ersten drei Feldern stehen die dem Affix zugeordnete Lautschrift, Information über die Wortart, soweit sie anhand des Affixes erkennbar sind, und eine Betonungsinformation zur Bestimmung des Wortakzents. Diese Daten müssen für alle Affixe vorhanden sein und werden vom Interpreter weitergeleitet. Die optionalen Felder enthalten beliebig viele Befehle an den Interpreter.

Lautschrift	Wki.	Bi.	Befehl 1			Befehl n	0
-------------	------	-----	----------	--	--	----------	---

Bild 5.1: Aufbau der den Affixen zugeordneten Verifizierungseinträge

Dabei bedeuten:

Lautschrift: die dem Morphem zuzuordnende Lautschrift

Wki. : Wortklassen-Information

Bi. : Betonungs-Information

Befehl 1 ... Befehl n: Befehle für den Interpreter,
n = 0,1,2, ...

0 : Endekennzeichnung

5.1.2 Befehle zur Verifizierung von Affixen

Von den o.a. drei Bedingungen, die zur Verifizierung eines Affix erfüllt sein müssen, wird nur die "Stammfähigkeit" des angrenzenden Restwortes immer überprüft; und zwar nachdem der Interpreter den Verifizierungseintrag abgearbeitet und festgestellt hat, daß ein Affix vorliegt. Die anderen beiden Bedingungen sind durch im Verifizierungseintrag gegebene Kommandos zu erfüllen. Dazu steht eine Reihe von Befehlen zu Verfügung.

Jeweils ein Befehl veranlaßt eine Überprüfung ; ob vor dem Affix ein Finalcluster bzw. nach dem Affix ein Initialcluster steht. Diese Befehle bestehen nur aus einem Kommando und brauchen keine Daten.

Weitere Befehle befassen sich mit Vergleichen von Buchstaben vor oder nach potentiellen Affixen. Neben der Angabe, ob vor oder nach dem Affix gesucht werden muß, benötigen diese Befehle auch Angaben über die Länge der zu vergleichenden Zeichenketten sowie die Zeichenketten selbst. Da sich gezeigt hat, daß zur Verifizierung der Vergleich von drei angrenzenden Buchstaben ausreicht, wurde die Angabe über die Länge des zu vergleichenden Strings mit in den Befehl einbezogen, dadurch ergeben sich insgesamt 6 Befehle, drei zur Suche vor Affixen und drei zur Suche hinter Affixen. Hinter dem Kommando können beliebig viele Zeichenfolgen mit der angegebenen Länge angeführt werden, die alle verglichen werden, wenn nicht vorher eine Übereinstimmung festgestellt wird. Um z.B. festzustellen, ob das Suffix 'in' Teil der Wörter 'Termin', 'Medizin' oder 'Vitamin' ist, muß hinter dem Kommando, das eine Überprüfung von drei Zeichen langen Folgen vor dem Suffix veranlaßt, die Zeichenkette 'ERM,DIZ,TAM,ende' übergeben werden. 'ende' markiert das Ende der Zeichenkette.

Nach Ausführung des Befehls hat der Interpreter entweder eine in den Daten angegebene Buchstabenfolge gefunden oder keine Übereinstimmung festgestellt. In beiden Fällen ist nicht bekannt, was das Ergebnis für die Verifizierung bedeutet. Deshalb ist jedem der sechs Befehle ein Modifizierer zugefügt. Dieser Modifizierer gibt an, was zu tun ist, wenn eine Zeichenfolge gefunden wurde, und er kann vier verschiedene Werte annehmen:

- 0 : wahrscheinliches Affix, Verifizierungseintrag weiter verarbeiten
- 1 : wahrscheinlich kein Affix, Verifizierungseintrag weiter verarbeiten
- 2 : Affix sicher erkannt, Verifizierung abbrechen
- 3 : es liegt kein Affix vor, Verifizierung abbrechen

Die Bedeutung der beiden letzten Werte liegt auf der Hand; doch zum Verständnis der Bedeutung der ersten Werte muß man die Funktionsweise des Interpreters kennen.

Vor Beginn der Ausführung des Verifizierungseintrags wird eine Bewertung der Verifizierung voreingestellt, und zwar auf den Mo-

difizierer '0', was bedeutet, daß die identifizierte Zeichenfolge als wahrscheinliches Affix angenommen werden soll. Sind im Verifizierungseintrag keine Befehle vorhanden, wird wegen des gültigen Modifizierers '0' entschieden, daß ein Affix vorliegt. Müssen Befehle ausgeführt werden, führt jeder erfolgreiche Vergleich dazu, daß der zugeordnete Modifizierer als Bewertung übernommen wird. Bei erfolglosen Vergleichen bleibt der alte Modifizierer erhalten. Werden die Modifizierer '2' oder '3' übernommen, bewirken sie einen Abbruch der Verifizierung. Bei den Modifizierern '0' und '1' werden die weiteren Befehle bearbeitet; und erst wenn das Ende des Eintrags erreicht ist, wird aufgrund des gültigen Modifizierers entschieden, ob das Affix verifiziert ist.

Der Ablauf der Verifizierung sei an einem vereinfachten Beispiel verdeutlicht: 'ge' soll in dem Wort 'Gebet' und immer dann, wenn es keine Form von 'geben' ist, als Vorsilbe identifiziert werden.

- Überprüfe, ob das folgende Restwort mit einem Initialkuster beginnt. Mit diesem Befehl wird nicht nur bei 'gebt', sondern auch bei einer ganzen Reihe weiterer Wörter wie 'gerne', 'Gerste', 'gelb' usw. festgestellt, daß keine Vorsilbe vorliegt, weil das Restwort nicht mit einem initialen Grundkuster beginnt. Die Verifizierung wird abgebrochen, und das Affix gilt als nicht verifiziert. Wird dagegen ein Initialkuster festgestellt, bleibt der voreingestellte Modifizierer '0' erhalten und die weiteren Verifizierungseinträge werden bearbeitet.
- Untersuche mit Modifizierer '2', ob die Zeichenfolge 'bet' hinter dem Präfix steht. Ist dies der Fall, wird bei dem Wort 'Gebet' die Verifizierung abgebrochen und 'ge' als Vorsilbe markiert. Folgt 'bet' nicht, hat der Befehl keine Auswirkung und der Modifizierer '0' bleibt erhalten.
- Untersuche mit Modifizierer '3'; ob die Zeichenfolge 'be' hinter dem Präfix steht. Mit diesem Befehl werden 'geben' und einige seiner Ableitungen erkannt. 'ge' ist in diesem Fall keine Vorsilbe und die Verifizierung wird mit negati-

vem Ergebnis abgebrochen. Folgt 'be' nicht, hat der Befehl keine Auswirkung und der Modifizierer '0' bleibt erhalten.

- Endekennzeichnung. Jetzt wird der gültige Modifizierer untersucht; und '0' besagt; daß die Vorsilbe als erkannt gelten soll. Sie wird markiert und die Verifizierung wird mit positivem Ergebnis abgebrochen.

Neben diesen Befehlen, die der reinen Verifizierung dienen, gibt es noch weitere Befehle, die bei erfolgreichem Vergleich auch die Lautschrift und/oder die Betonungsinformation ändern können oder die einen oder mehrere Buchstaben zum Affix hinzuziehen. Sie behandeln Sonderfälle; wie z.B. die Unterscheidung der Suffixe 'al' und 'sal'; 'es' und 'oes'. Diese Befehle benötigen keinen Modifizierer, weil sie schon eine erfolgreiche Verifizierung voraussetzen. Deshalb wird nach solchen Änderungen die Verifizierung immer abgebrochen.

Eine vollständige Liste der Interpreter-Kommandos findet sich in Anhang 4.

5.1.3 Die Erkennung von fremdstämmigen betonten Endungen

Das Verfahren zur Wortanalyse ist auf deutschstämmige Wörter abgestimmt. Die Klusteranalyse kann Fremdwörter zwar nicht korrekt zerlegen, aber sie zerstört sie auch nicht so, daß dies häufig schwerwiegende Phonemisierungsfehler zur Folge hätte.

Einer der häufigsten Aussprachefehler bei Fremdwörtern; die nach deutschen Regeln gesprochen werden; ist ein falsch gesetzter Hauptakzent, denn deutsche Wörter werden meist auf der ersten Silbe des Stammes betont, Fremdwörter dagegen auf einer der letzten Silben. Solche typischen stamffinalen Silben; die immer den Hauptakzent tragen; sind z.B. 'abel'; 'age'; 'og'; 'aille' und 'thek'. Neben den immer betonten stamffinalen Silben gibt es noch weitere, die nur unter bestimmten Bedingungen betont werden wie 'or' und 'ik'.

Zur Bestimmung dieser fremdstämmigen Silben bietet sich die Affixerkennung an. Es handelt sich bei den stamfinalen Silben zwar nicht um Suffixe im linguistischen Sinn; aber praktisch können sie genau wie Suffixe behandelt werden.

Deshalb wurden Untersuchungen an einem auf Datenträger vorliegenden Lexikon der deutschen Sprache vorgenommen (Mackensen, 1962). Sämtliche Wörter wurden untersucht; in denen in finaler Stellung eine Buchstabenfolge auftrat; die einer der von Siebs (1969) angegebenen fremdstämmigen Endungen entsprach. Für viele Buchstabenfolgen wurden Regeln aufgestellt; mit deren Hilfe fremdstämmige stamfinale Silben verifiziert und akzentuiert werden können. Leider konnten für einige Silben, die mit deutschstämmigen Flexionsendungen und Affixen zusammenfallen wie 'el' und 'en', keine Regeln gefunden werden.

Untersuchungen am in Kap. 4.4 beschriebenen Testkorpus haben gezeigt, daß die von Siebs angegebenen finalen Silben nur einen kleinen Teil der Silben ausmachen, die für eine korrekte Akzentuierung der fremdstämmigen Wörter in den Listen erfaßt werden müssen. An einer Erweiterung der Suffixlisten und der Erstellung des zugehörigen Regelsatzes wird gearbeitet.

5.1.4 Fehler bei der Erkennung von Affixen

Bei der Erkennung von Affixen sind drei verschiedene Fehlertypen möglich: noch nicht berücksichtigte Affixe, Nicht-Erkennungen durch Fehler der Clusteranalyse und falsche Abtrennungen.

Bei den Präfixen traten Fehler weitaus am häufigsten durch noch nicht in den Listen enthaltene Vorsilben auf, und zwar aus zwei Gründen: In den Listen sind nur sehr wenige fremdstämmige Vorsilben enthalten, weil sie für die Betonung nicht von Bedeutung sind und nur zu wenigen Aussprachefehlern führen. Die meisten im Textkorpus nach Kap. 4.4 auftretenden 38/139 fremdstämmigen Vorsilben konnten also nicht erkannt werden. Andererseits sind die meisten deutschstämmigen Vorsilben zwar in der Liste enthalten, aber es gibt eine Reihe von Morphemen wie 'außer', 'inner', 'einer' und

'aller', die keine eigentlichen Vorsilben, jedoch wie solche zu behandeln sind. Diese uneigentlichen Vorsilben, die im Korpus 13/35 mal auftreten, sind bisher noch nicht in den Listen enthalten. Nicht in den Listen enthaltene Vorsilben führten nur in 5/14 Fällen zu schweren Aussprachefehlern.

In 5/13 Fällen wurden Vorsilben durch einen Fehler der Clusteranalyse nicht erkannt. Diese Fehler führen in allen Fällen zu einer Minderung der Verständlichkeit.

Falsch erkannte Vorsilben sind dagegen weit weniger häufig als nicht erkannte. Nur in 8/29 Fällen wurden Teile eines Fremdwortes als deutsche Vorsilbe abgetrennt wie z.B. bei den Wörtern 'Generation' und 'sozial', daraus resultierten 2/4 Aussprachefehler. Fehler dieser Art können durch Erweiterung der Verifizierungsregeln völlig abgestellt werden.

Bei den Nachsilben traten die gleichen Fehlertypen auf. Bedingt durch die noch nicht vollständige Implementation des Regelsatzes für fremdstämmige Wortteile werden 40/181 fremdstämmige finale Wortteile noch nicht erkannt, was in den meisten Fällen zu Betonungsfehlern führt.

Ein Programmfehler, der Ergebnisse der Clusteranalyse falsch interpretiert, führte in 11/36 Fällen zur Nichterkennung einer Nachsilbe. Dies erzeugt in 4/15 Fällen einen schweren Aussprachefehler. In diesen Fällen wird ein Fugen-'n' oder 'l', das durch die Auslöschung des 'e' in 'en' oder 'el' entstanden ist, nicht als Fuge erkannt, sondern als Reststamm interpretiert; und zwar in Wörtern wie 'Begegnung' oder 'neblig'. Dies führt bei der Überprüfung, ob der Reststamm stammfähig ist, zu einem negativen Ergebnis.

Nur in 1/3 Fällen, nämlich bei dem Wort 'Aussage'; wurde ein Teil eines deutschen Wortes als fremdstämmige Endung erkannt.

Aus der großen Zahl nicht erkannter Affixe wird deutlich; daß die Listen der Affixe bei weitem noch nicht vollständig sind. Die wenigen weiteren Fehler zeigen jedoch, daß das Prinzip der Affix-

erkennung mit einer Verifizierung von Affixen anhand des Umfelds und anhand des Wortaufbaus nicht nur zur Bestimmung von Affixen im eigentlichen Sinne, sondern auch zur Bestimmung von finalen Wortteilen in Fremdwörtern sehr gut geeignet ist.

5.2 Die Erkennung von Endungen

Ehe von der Affixerkennung die am Wortende stehenden Suffixe erkannt werden können, müssen die noch hinter den Suffixen stehenden Endungen abgetrennt werden. Unter dem Begriff 'Endungen' sind dabei nicht nur die eigentlichen Flexionen, d.h. Deklinations- und Konjugationsendungen zusammengefaßt, sondern auch die Komparationsendungen von Adjektiven ('-er' und '-est') sowie einige Schwa-Endungen ('-e', 'el', '-er', '-en' und '-chen'), die auch als Affix oder als Teil des Stamm-Morphems auftreten können. Wie an dem in Kapitel 4 beschriebenen Wortaufbau postuliert wurde, können Endungen nur am Wortende auftreten. Die Schwa-Endungen, die auch in Wortzusammensetzungen am Ende eines Teilwortes auftreten können, werden zusätzlich als Suffix behandelt. Durch dieses Vorgehen ist ein fester Ausgangspunkt für die Suche nach Affixen am Wortende vorgegeben, was die Suche vereinfacht und die Ergebnisse einfacher überprüfbar macht.

Es bietet sich an, für die Abtrennung von Endungen ein der Suffixerkennung ähnliches, auf Listen basierendes Suchverfahren zu verwenden. Dies wurde zunächst versucht, wobei die Endungsliste sämtliche von Kästner (1972) angegebenen Schwa-Endungen enthielt. Dabei wurde unter nur geringer Berücksichtigung des grammatikalischen Aufbaus des Wortes eine möglichst große zusammenhängende Gruppe dieser Endungen am Wortende abgetrennt. Von vornherein wurde dabei mit der Verstümmelung von Stämmen in der Art von

'Lavafelsen' → 'Lavaf-el-s-en'

gerechnet. Es traten jedoch mehr Fehler auf als ursprünglich angenommen, so daß die Verständlichkeit unzumutbar beeinträchtigt wurde.

Deshalb wurde aus den zunächst verwendeten Schwa-Endungen unter Berücksichtigung aller möglichen Konjugationen, Deklinationen, Steigerungen und Schwa-Endungen eine Grammatik entwickelt, die angibt, welche Buchstabenfolgen als Wortendung zulässig sind. Den Aufbau dieser Grammatik zeigt Anhang 5.1. Wie man sieht, lassen sich nur wenige Buchstaben bzw. Buchstabenfolgen zu ca. 200 Endungsfolgen kombinieren.

Der Algorithmus zur Erkennung und Verifizierung von Affixen ist in der beschriebenen Form zur Abtrennung von Endungen nicht geeignet, weil er eine Verifizierung über die Grammatik nicht durchführen kann. Deshalb wurde zur Identifizierung von Endungen ein spezielles Programm geschrieben. Dieses Programm führt auf drei Ebenen drei verschiedene Aufgaben durch: Ausführung der Grammatik auf der höchsten Ebene, Identifizierung von Endungsteilen als Buchstabenfolge und Verifizierung von Buchstabenfolgen als Endungsbestandteil auf der niedrigsten Ebene.

Auf der mittleren Ebene erkennt eine Reihe von Identifizierungsroutinen, vom Wortende oder von der letzten abgetrennten Endung her beginnend, jeweils eine Zeichenfolge als Mitglied einer Endungsklasse. Die Endungsklassen treten in der Grammatik in Anhang 5.1 als durch geschweifte Klammern zusammengefaßte Gruppen oder als Untermengen dieser Gruppen auf. Z.B. bilden die Zeichenfolgen '(n)end', 'elnd' und 'ernd' die Endungsklasse des Partizip Präsens, und diese Endungsklasse kann mit der Klasse der stammfinalen Schwa-Endungen, die 'en' und 'er' enthält, zu einer übergeordneten Klasse zusammengefaßt werden.

Nachdem eine Endung identifiziert wurde, ruft die Identifizierung eine Verifizierungsroutine auf, die abhängig ist vom identifizierten Endungstyp oder vom Anfangsbuchstaben der Endung. Für eine vom Endungstyp abhängige Verifizierung sind in Anhang 5.2 die Regeln zur Bestätigung der Verkleinerungsendung 'chen' als Beispiel angegeben, für eine vom Anfangsbuchstaben abhängige Verifizierung stehen in Anhang 5.3 die Regeln für mit 't' beginnende Endungen.

Neben der Bestätigung von Endungen durch das Buchstabenumfeld, wie sie von den Verifizierungsroutinen auf der untersten Ebene durchgeführt wird, müssen die Endungsfolgen den Regeln der Grammatik entsprechen. Für die Einhaltung der Grammatik wird auf der höchsten Ebene gesorgt. Ausgehend vom letzten, noch nicht einer Endung zugeordneten Buchstaben wird eine Vorauswahl getroffen, welche Endungszeichenfolgen zur Endung gezogen werden können. Dabei werden nur von der Grammatik erlaubte Folgen berücksichtigt. Nur für die noch verbleibende Menge von möglichen Zeichenfolgen werden die Identifizierungsroutinen aufgerufen.

Nachdem die längstmögliche Endung gefunden wurde, wird wie bei der Erkennung von Affixen überprüft, ob das verbleibende Restwort noch ein Stamm sein kann. Dazu wird das schon bei der Affixerkennung beschriebene Verfahren benutzt. Stellt die Routine jedoch fest, daß das Restwort nicht mehr stammfähig ist, wird nicht die gesamte Endung für ungültig erklärt, sondern nur die als letztes gefundene Zeichenfolge. Danach wird das Restwort erneut überprüft.

Daß Endungen nicht über Listen, sondern von Programmteilen erkannt werden, macht das Programm etwas unflexibler, führt aber zu einem weit schnelleren Ablauf und schafft dadurch Zeitreserven im Hinblick auf einen vollständigen Programmausbau bei Echtzeitbetrieb.

Die Endungserkennung wurde anhand der o.a. 1114/8533 Wörter getestet, dabei führte eine falsche Endungsabtrennung bei 20/67 Wörtern zu Aussprachefehlern; was auf 768/3049 Wörter mit Endungen bezogen einer Fehlerquote von 2,6/2,2 % entspricht. Von diesen Wörtern waren 11/49 fremdstämmig wie z.B. 'Hotel'; 'parallel' und 'System', entsprachen also in Aufbau und Aussprache nicht den auf deutsche Wörter abgestimmten Regeln. In deutschstämmigen Wörtern treten hauptsächlich Fehler dadurch auf; daß Stämme oder Teile des Stamms falsch zur Endung gezogen werden wie in 'außerdem' und 'Wochenende'; und daß oft zweifelhaft ist, ob ein 't' zum Stamm gehört wie z.B. in dem Gegensatzpaar

'Kante' - 'geplante'.

5.3 Die Erkennung von Ausnahmen

Für die Satzanalyse ist es wichtig, daß die sog. Formwörter in ihrer Wortart identifiziert werden. Dies ist nach Regeln nicht möglich, sondern die Identifizierung muß über Listen durchgeführt werden. Neben den Wörtern, die zwar korrekt umgesetzt werden, aber für die Satzanalyse identifiziert werden müssen, können auch Wörter, die fehlerhaft umgesetzt werden, über die Ausnahmenerkennung identifiziert und korrekt umgesetzt werden, wenn sie häufig genug auftreten, um für die Fehlerquote von Belang zu sein. Für viele sehr häufige Formwörter wie z.B. 'bis', 'mit', 'das', 'hat' ist beides der Fall, denn in Formwörtern werden Vokale entgegen den allgemeinen Ausspracheregeln oft gekürzt. Sinnwörter sind dagegen weitaus stärker an den allgemeinen Ausspracheregeln orientiert.

Viele Formwörter können nicht flektiert werden, haben also keine Endung. Außerdem ist es möglich, daß Wörter durch eine falsch abgetrennte Endung falsch ausgesprochen werden. Diese beiden Fälle werden von einer Ausnahmenerkennung berücksichtigt, die Wörter untersucht, ehe eine Endung abgetrennt wird. In den Listen dieser Ganzwort-Ausnahmenerkennung gibt es bisher 172 Eintragungen. Dabei handelt es sich nur um Formwörter, falsch umgesetzte Sinnwörter sind in dieser Liste noch nicht enthalten.

Einige Formwörter und alle Sinnwörter treten flektiert auf. Da durch die Flexion die Anzahl der Formen vervielfacht wird, ist es bei flektierbaren Wörtern sinnvoll, nicht alle Formen abzuspeichern, sondern nur Stämme. Diese Stamm-Ausnahmen werden von einer zweiten Ausnahmenerkennung untersucht, die bis auf die Tatsache, daß sie mit anderen Listen arbeitet, mit der Ganzwort-Erkennung identisch ist. Diese Stamm-Erkennung bearbeitet keine Stämme im linguistischen Sinn, sondern die Teilwörter, die nach Abtrennung der Endung verbleiben. Wenn die Flexion z.B. zu einer Vokaländerung im Stamm führt, bleibt das unberücksichtigt. Deshalb müssen z.B. zu dem Hilfsverb 'können' die drei Stämme 'koenn', 'konn' und 'kann' und zu 'haben' 'hab', 'hatt', 'haett' und die Ganzwort-Ausnahmen 'hat' und 'hast' abgespeichert werden. Die Liste dieser Stamm-Ausnahmen mit 44 Eintragungen enthält derzeit nur

Formwörter.

5.4 Zusammenfassung

Da die Klusteranalyse allein keine vollständige morphologische Zerlegung von Wörtern durchführen kann, wird sie von drei Verfahren ergänzt: Ausnahmenerkennung, Endungserkennung und Affixerkennung.

Die Ausnahmenerkennung wird zur Erkennung von ganzen Wörtern und zur Erkennung von Wörtern mit abgetrennter Endung benötigt. Sie bestimmt die Wortart der für die Satzbetonung wichtigen Formwörter.

Die Endungserkennung bestimmt und verifiziert die nur am Wortende auftretenden Konjugations-, Deklinations- und Komparationsendungen, wobei unter Beachtung einer Grammatik eine möglichst große Folge von Endungsbestandteilen zu einer Gesamt-Endung zusammengefaßt wird.

Zur Erkennung von Affixen wurde ein Verfahren entwickelt, das anhand von Listen Affixe als Buchstabenfolge identifiziert und mit ebenfalls in den Listen enthaltener Information die Zeichenketten als Affixe verifiziert. Dies geschieht über Vergleiche der angrenzenden Buchstaben des Restwortes und durch allgemeine Untersuchungen des Aufbaus des Restwortes. Das Verfahren erkennt und verifiziert ebenfalls zu betonende stamminale Silben in Fremdwörtern.

6. DIE UMWANDLUNG VON WÖRTERN

In den beiden letzten Kapiteln wurden Verfahren vorgestellt, mit denen deutsche Wörter mit einer für die Sprachsynthese ausreichenden Genauigkeit analysiert und in Morpheme zerlegt werden können. Durch die Verwendung dieser Verfahren vereinfachen sich die Regeln zur Erzeugung von Lautschrift und Wortakzent im Vergleich zu Systemen mit Umwandlung auf Wortebene stark. Die Umwandlung kann mit ähnlich einfachen Regeln wie den im Duden (Mangold, 1974a) angegebenen durchgeführt werden. Weil die benötigten Informationen vorab von der Wortanalyse geliefert werden, können die Verfahren zur Phonemisierung und zur Bestimmung des Wortakzents unabhängig voneinander arbeiten, während sie bei der Umsetzung auf Wortebene zusammenarbeiten müssen, wenn Doppelanalysen vermieden werden sollen.

Um Lautschrift und Akzent darzustellen, wird in der Linguistik die international gebräuchliche IPA-Lautschrift verwendet. Die IPA-Lautschrift benutzt jedoch viele mit üblichen Druckern nicht darstellbare Sonderzeichen und ist deshalb nur für Satzdruck geeignet. Im englischen Sprachraum für die Sprachsynthese entwickelte Lautschriften sind dagegen zwar mit üblichen Zeichensätzen auf dem Terminal darstellbar, aber an englischen Ausspracheregeln orientiert und daher für die deutsche Sprache schwer lesbar. Deshalb wurde zur Kontrolle der Umsetzung eine eigene an Aufgabe und deutsche Sprache angepaßte Lautschrift entwickelt. Diese Lautschrift wird auch in den folgenden Abschnitten bei der Darstellung der Transkriptionsregeln verwendet. In Anhang 6 ist der Zusammenhang zwischen der benutzten Lautschrift und der IPA-Lautschrift dargestellt.

6.1 Die Umwandlung von Text in Lautschrift

Weil die Wörter in Morpheme zerlegt vorliegen und weil ihr Aufbau aus Stämmen, Vorsilben, Nachsilben und Endungen explizit bekannt ist, lassen sich zur Umwandlung der Wörter in Lautschrift

sehr einfache Regeln angeben. Diese Regeln lassen sich in zwei Klassen einteilen:

- Morphembezogene Regeln überführen Grapheme oder Graphemfolgen innerhalb von Morphemen in Phoneme oder Phonemfolgen; wie es z.B. die Regel 'OH' → 'o' tut. Die Regeln können dabei vom Morphemtyp abhängig sein, z.B. gilt die Regel 'E' → 'š' nur in Endungen.
- Regeln, die über Morphemgrenzen hinweg wirken, überführen ebenfalls Grapheme oder Graphemfolgen in Phoneme oder Phonemfolgen, wobei aber nicht nur das Umfeld innerhalb des Morphems, sondern auch der Typ und das Umfeld des angrenzenden Morphems berücksichtigt werden. So gilt z.B. für ein 'B' am Morphemende, daß es in ein stimmloses 'P' überführt wird, wenn es am Wortende steht, das nachfolgende Morphem ein Stamm, eine Fuge oder eine Vorsilbe ist, oder die folgende Nachsilbe oder Endung mit einem Konsonanten beginnt.

Werden Morpheme über die Ausnahmenlisten phonemisiert, können nur die morphembezogenen Regeln berücksichtigt werden. Nach Listen können Anpassungen an den Morphemgrenzen nach grenzübergreifenden Regeln nur durchgeführt werden, wenn sie durch zusätzliche Befehle im Regelsatz berücksichtigt werden. Um diese Schwierigkeit zu umgehen, wurde die Umwandlung von Text in Lautschrift in zwei Schritte aufgeteilt. Im ersten; Phonemisierung genannten Schritt werden alle Morpheme, die nicht anhand der Ausnahmenliste in Phoneme überführt worden sind, ohne Berücksichtigung der grenzübergreifenden Regeln in eine lautschriftliche Beschreibung umgewandelt, wie sie auch in den Ausnahmenlisten verwendet wird. Diese lautschriftliche Beschreibung wird dann in einem zweiten Schritt von der sogenannten morphembezogenen Koartikulation anhand der grenzübergreifenden Regeln korrigiert. Grenzübergreifende Regeln wandeln jetzt allerdings nicht mehr Grapheme in Phoneme um, sondern sie überführen Phoneme in Abhängigkeit von Umfeldbedingungen in andere Phoneme oder Allophone. Diese Art der Transformation wird Koartikulation genannt.

Die Koartikulation ist morphembezogen, weil sie nur über Morphemgrenzen und nicht über Wortgrenzen hinweg wirkt. Koartikulationen über Wortgrenzen können nicht auf der Wortebene, sondern erst auf der Satzebene durchgeführt werden, denn auf der Wortebene sind die Nachbarwörter nicht bekannt.

Die Koartikulationsregeln können zwar über Morphemgrenzen hinweg wirken, müssen es aber nicht; sie ändern die Aussprache bestimmter Laute im Morphem genauso wie an den Morphemgrenzen.

6.1.1 Die Darstellung der Transkriptionsregeln

Zur Darstellung der Umsetzungsregeln für Phonemisierung und Koartikulation wurde folgendes Format gewählt:

(Linkes Umfeld) + Zeichenfolge + (rechtes Umfeld) ->
Phonemfolge 1 sonst Phonemfolge 2

Durch das linke und rechte Umfeld sind zwei Bedingungen gegeben, die erfüllt sein müssen, damit die Zeichenfolge durch die Phonemfolge 1 ersetzt wird. Sind sie nicht erfüllt, wird die Zeichenfolge durch die Phonemfolge 2 ersetzt. Bedingungen lauten z.B.:

"Konsonant" : Zeichenfolge grenzt direkt an einen Konsonanten
"Vokal" : Zeichenfolge grenzt direkt an einen Vokal
"Morphemgrenze": Zeichenfolge grenzt direkt an ein anderes Morphem
"Endung" : Zeichenfolge grenzt direkt an eine Endung.

Morphemgrenzen anderer Morpheme werden wie die von Endungsmorphemen durch Angabe des Morphemnamens dargestellt. Linkes und rechtes Umfeld sind optional: werden sie nicht angegeben, wird auch das zugehörige Pluszeichen weggelassen. In einem Umfeld können mehrere Bedingungen angegeben werden, von denen nur eine erfüllt sein muß. Die Bedingungen werden durch '!' getrennt. Dann bedeutet z.B.:

(B!D!Vokal) + ...

daß links der Zeichenfolge ein 'B' ein 'D' oder ein Vokal stehen muß. Wenn gleiche Regeln für mehrere Zeichenfolgen gelten, können die Zeichenfolgen und die Phonemfolgen analog angegeben werden, wobei der n-ten Zeichenfolge auch die n-te Phonemfolge zugeordnet ist, wie z.B. in

B!D!G + (Konsonant) → P!T!K sonst B!D!G

Die Phonemzeichenfolgen werden in der in Anhang 6 vorgestellten Lautschrift notiert. Zeichenfolge und explizit angegebene Zeichen im Umfeld werden im Zeichensatz der jeweils gültigen Darstellungsebene notiert, d.h. für Phonemisierungsregeln werden Grapheme angegeben und für Koartikulationsregeln Phoneme. Alle Regeln für umzusetzende Zeichenfolgen mit gleichem Anfangszeichen sind zu einer Gruppe zusammengefaßt. Innerhalb der Gruppe herrscht eine Hierarchie derart, daß innerhalb einer Gruppe die Regeln von oben nach unten abgearbeitet werden. Sobald eine Regel zutrifft, kann die entsprechende Zeichenfolge ersetzt werden, die nachfolgenden Regeln brauchen nicht mehr bearbeitet zu werden. Die Regeln nehmen deshalb von oben nach unten immer allgemeineren Charakter an, und die letzte ist eine Auffangregel, die alles umsetzt, was übrig bleibt.

6.1.2 Die Phonemisierung von Vokalen

Außer den Diphthongen und dem Schwa können alle deutschen Vokale entweder lang oder kurz realisiert werden. Da beide Ausführungsformen, von Ausnahmen abgesehen, im Schriftbild durch die gleichen Vokalgrapheme codiert werden; und die Regeln für die Entscheidung 'kurz-lang' bis auf wenige Ausnahmen für alle Vokale gleich sind, wurde der Phonemisierungsvorgang für Vokale in zwei Schritte zerlegt. Im ersten Schritt wird anhand der Vokalgrapheme eine sog. Vokalqualität bestimmt, die bezüglich der Vokallänge noch nicht festgelegt ist, und im zweiten Schritt wird anhand der für Vokale gemeinsamen Umfeldregeln die 'kurz-lang'-Entscheidung getroffen. Ausgenommen davon sind die Diphthonge und das Schwa,

für die keine Opposition 'kurz-lang' besteht, und Vokale, deren Länge durch besondere Vokalgrapheme oder durch besondere Umfeldregeln bereits im ersten Schritt festgelegt werden kann.

Es folgen nun die Regeln der ersten Phonemisierungsstufe. Zur Kennzeichnung der noch in der Länge unbestimmten Vokalphoneme wurde der entsprechende kurze Vokal mit nachgestelltem Minus-Zeichen verwendet.

AEU -> O/

AI -> E/

AY -> E/

AU -> A/

AE -> V-

A -> A-

EI -> E/

EY -> E/

EU -> O/

E + (H) -> e

EE -> e

E + (im Endungsmorphem) -> \$

E + (E erster Vokal im Morphem) -> E- sonst E

IE -> i

I -> I-

OE -> Q-

O -> O-

UE + (CH) -> y

UE -> Y-

U + (CH) -> u

U -> U-

Y + (Morphemgrenze) -> I-

Y -> Y-

Zu den Regeln für das 'y' ist zu bemerken; daß hier nur die Regeln für vokalisches 'y' aufgeführt wurden; eine weitere Regel für konsonantisches 'y' findet sich bei den Umsetzungsregeln für Konsonanten. Die Entscheidung darüber, ob 'y' konsonantischen oder vokalischen Charakter hat, ist schon in einem früheren Programmteil getroffen worden, als für die Klusteranalyse Konsonanten- und Vokalkluster markiert wurden.

Im zweiten Schritt wird für die noch teilweise unbestimmten Vokale die Länge ermittelt. Für den zu untersuchenden Vokal sei

V_g die zugehörige Graphemfolge
 v der zugehörige lange Vokal
 V der zugehörige kurze Vokal.

Dann gilt:

$$\begin{array}{lll} V_g + (H! \text{Morphemgrenze}) & \rightarrow & v \\ & V_g V_g & \rightarrow & v \\ & V_g + (P!X) & \rightarrow & V \\ V_g + (2 \text{ oder mehr Konsonanten}) & \rightarrow & V \text{ sonst } v \end{array}$$

Die Durchführung der Phonemisierung von Vokalen in zwei Schritten führt zu einer Vereinfachung des Regelsatzes. Im ersten Schritt liefern zehn Regeln in der Länge unbestimmte Vokale; die im zweiten Schritt mit vier Regeln bearbeitet werden. Bei einer einstufigen Umsetzung würden multiplikativ aus 14 Regeln 40 werden, d.h. der Umfang des gesamten Regelsatzes würde sich verdoppeln.

6.1.3 Die Phonemisierung von Konsonanten

Die meisten Konsonantengrapheme lassen sich aufgrund des gewählten Umsetzungsverfahrens sehr einfach in Konsonantenphoneme überführen. Viele lassen sich mit einer einzigen Regel phonemisieren, wenn man für alle Konsonanten sicherstellt, daß die folgende Zusatzregel erfüllt wird:

kk $\rightarrow$ p

Diese Regel besagt, daß zwei aufeinanderfolgende gleiche Konsonantengrapheme 'k' in nur ein Konsonantenphonem 'p' überführt werden. 'p' ist dabei das Phonem, das aufgrund der folgenden Regeln dem ersten der beiden Konsonantengrapheme zugeordnet wird.

Mit dieser Zusatzregel werden folgende Konsonanten durch nur eine einzige Regel eindeutig phonemisiert.

Z -> T/
 X -> K/
 W -> W
 QU -> KW
 N -> N
 M -> M
 L -> L
 K -> K
 J -> J
 G -> G
 F -> F
 D -> D
 B -> B
 Y -> J

Auch für die Umsetzung der restlichen Konsonanten außer 'C' werden nur maximal drei Regeln benötigt:

H + (Vokal) -> H sonst '

 PH -> F
 P -> P

 RH -> r
 R -> r

 SCH -> S/
 (Stamm) + S + (T!P) -> S/
 S + (Morphemgrenze!Vokal) -> Z sonst S

T + (Z) -> '
 TH -> T
 T -> T

(Morphemgrenze!Konsonant) + V -> F
 V + (Vokal) -> W sonst F

Nur für das 'C' ist der Regelsatz umfangreicher. Dies ist darauf zurückzuführen, daß es in deutschstämmigen Wörtern als Teil des 'CH' oder des 'CK' und in fremdstämmigen Wörtern alleinstehend oder als Teil des 'CH' mit verschiedenen Aussprachen auftreten kann.

(Morphemgrenze) + CH + (A!O!U!Konsonant) -> K
 (Morphemgrenze) + CH + (Vokal) -> C
 CH + (S, aber nicht ST) -> K
 (EU) + CH -> C
 (A!O!U!) + CH -> X sonst C
 CC -> K
 CK -> K
 C + (AE!E!I!Y!OE!) -> T/ sonst K

Die angegebenen insgesamt 36 Regeln liefern eine Lautschrift, die noch fehlerhaft ist, denn es wurden bisher nur die morphembezogenen Regeln angewendet. Diese Fehler werden im folgenden Programmteil von den Koartikulationsregeln korrigiert.

6.1.4 Koartikulation

Die Koartikulationsregeln behandeln nur bestimmte Konsonanten, und sie lassen sich in zwei Klassen einteilen: Regeln zur Behandlung von 'N' bzw. 'N/' und Regeln zur Entstimmung, d.h. zur Stimmlos-Umschaltung bei stimmhaften Frikativen und Plosiven bzw. zur Allophonwahl zwischen 'r' und 'R'.

Die erste Gruppe besteht nur aus zwei Regeln:

N + (G!K!) -> N/
 (N/) + G + (Morphemgrenze!Konsonant!\$) -> '

Mit der ersten Regel ist sowohl die bewußte Aussprache als 'N/' wie in 'Dinge' als auch die unbewußte; aber in der Hochlautung übliche Assimilation des 'N' an Morphemgrenzen wie in 'Weingebiet' erfaßt. Die zweite Regel ist eigentlich eine morphembezogene, also nicht grenzübergreifende Regel; sie kann aber erst jetzt angewandt werden, weil sie auf der ersten Regel aufbaut. Sie löscht das folgende 'G' in Wörtern wie 'dinglich'; 'Pfingsten' und 'Dinge', nicht aber in Wörtern wie 'ungerade', 'Ungarn' oder 'Bingo'.

Die zweite Gruppe enthält zwei Regeln zur Entstimmung von stimmhaften Frikativen bzw. zur 'r'-'R'-Festlegung und vier Regeln zur Entstimmung bei stimmhaften Plosiven.

Z!W!r!R + (Stamm!Vorsilbe!Fuge!Wortgrenze) -> S!F!R!R
 Z!W!r!R + (Vokal) -> Z!W!r!r sonst S!F!R!R

B!D!G + (Stamm!Vorsilbe!Fuge!Wortgrenze) -> P!T!K
 B!D!G + (Vokal am Anfang von Nachsilbe oder Endung) -> B!D!G
 B!D!G + (Nachsilbe!Endung) -> P!T!K
 B!D!G + (S!Z!T!D) -> P!T!K sonst B!D!G

Die 'R'-Allophone nehmen unter den behandelten Zeichen eine Sonderstellung ein, weil sie auch auf der linken Seite der Regeln mit beiden allophonischen Varianten auftreten. Dies ist für die Phonemisierung, die nur ein 'r' erzeugt, nicht notwendig, es hat sich aber gezeigt, daß auch geübten Mitarbeitern bei der Erstellung von Lautschrift für die Ausnahmenlisten Fehler unterlaufen, weil die beiden R-Laute rein allophonischen Charakter haben, während stimmhafte und stimmlose Frikative und Plosive unterschiedliche Phoneme sind. Durch die Behandlung beider Allophone werden also Lautschriftfehler in den Listen automatisch korrigiert, was das Aufstellen von Listen vereinfacht.

6.1.5 Der Knacklaut

Die bisher vorgestellten Regeln haben einen Laut noch nicht erzeugt: den Knacklaut. Der Knacklaut (engl.: glottal stop) ist ein Laut, der in der Schriftform der Sprache nicht codiert ist, sondern er wird nur bei der Aussprache an bestimmten Stellen vor Vokalen eingefügt. Dies geschieht dann, wenn akzentuierte Vokale am Anfang einer Vorsilbe oder eines Stammes stehen, z.B. vor dem 'A' in Beamter. Der Knacklaut braucht deshalb nicht von der Transkription erzeugt zu werden, sondern wird bei der Übergabe der transkribierten und akzentuierten Wörter an die Programme auf Satzebene automatisch eingefügt.

6.1.6 Fehler der Transkription

Zur Analyse von Fehlern der Phonemisierung und Koartikulation wurden wiederum die 1114/8882 zwei- oder mehrmals vorkommenden Wörter des schon in Kap. 4.4 beschriebenen Korpus gewählt.

Insgesamt wurden 148/675 Transkriptionsfehler gezählt. Über die Hälfte davon war von der Art, die nicht vorgewarnten Hörern nicht auffällt, z.B. das Weglassen oder Einfügen von Knacklauten bei falschen Zerlegungen oder aber das Ersetzen des Schwa durch ein kurzes 'E' aus dem gleichen Grund. Weiter waren einige zehn beabsichtigte, aber nicht verständlichkeitsmindernde Fehler vorhanden, z.B. wird 'st' am Teilwortanfang in Zusammensetzungen wie 'Lochstreifen' nicht als 'S/T' ausgesprochen.

Verständlichkeitsmindernde Fehler wurden insgesamt 54/282 gezählt, was einer Fehlerquote von 4,8/3,1 % entspricht. Davon entfielen 18/76 auf deutschstämmige Wörter und 36/206 auf Fremdwörter. Bei 220/762 im Korpus enthaltenen Fremdwörtern ergeben sich dadurch Fehlerquoten von 2/0,9 % für deutschstämmige Wörter und 16,4/27,0 % für Fremdwörter.

In deutschstämmigen Wörtern traten nur Fehler bei Vokalen auf, bis auf zwei Ausnahmen: In dem Wort 'Pflegeverein' wurde aufgrund einer nicht durchgeführten Zerlegung das Graphem 'V' in das Pho-

nem 'W' umgesetzt, und in dem Wort 'Aussage' wurde '-age' falsch als fremdstämmige Nachsilbe erkannt und wie in 'Passage' über Listen transkribiert. Die Transkription von Konsonanten ist also für deutschstämmige Wörter fast als fehlerfrei zu betrachten.

Bei den Vokalen in deutschstämmigen Wörtern wird das, was bei der Phonemisierung als 'Lautqualität' bezeichnet wurde, ebenfalls fast immer einwandfrei bestimmt, bis auf eine Ausnahme: In 'Wochenende' wurde das 'E' am Anfang von 'Ende' in ein Schwa umgewandelt, weil 'ende' falsch als Endung des Partizip Präsens erkannt wurde. Die restlichen Fehler bei den Vokalen lagen darin begründet, daß die Vokallänge falsch bestimmt wurde. Bei der Untersuchung der Ursachen für die falsche Entscheidung zeigten sich drei Fehlerquellen:

- Fehler der grammatikalischen Zerlegung. Hierzu gehörten sowohl nicht erkannte Morphemgrenzen als auch fehlerhaft erkannte Vorsilben, falsch abgetrennte und nicht abgetrennte Endungen in insgesamt 7/23 Wörtern.
- Fehler durch falsche Transkription der Vokallänge vor 'CH'. Es machte sich besonders das Wort 'Sprache' mit seinen Ableitungen bemerkbar, bei dem 4/25 mal das 'A' kurz transkribiert wurde und das so erheblichen Anteil an der Fehlerquote hatte.
- Fehler durch falsche Transkription der Vokallänge vor 'ß' bzw. 'SS'. Es machte sich störend bemerkbar; daß das 'ß' intern durch die Grapheme 'SS' dargestellt wird, so daß keine besonderen Regeln für die Vokallänge vor 'ß' eingeführt werden können. Während vor 'SS' immer kurze Vokale stehen, treten vor 'ß' auch oft lange auf.

An der niedrigen Fehlerquote von 2/0,9 % zeigt sich, daß die Transkription bis auf die beiden letztgenannten Fehlerquellen sehr gut an deutschstämmige Wörter angepaßt ist.

Für Fremdwörter wurde die hohe Fehlerquote von 16,4/27,0 % gemessen. Sie ist größtenteils darauf zurückzuführen, daß bisher nur ein Teil der Suffixe und noch gar keine Präfixe von fremdstämmigen Wörtern erkannt werden. So führte z.B. das nicht als Präfix gekennzeichnete 'am-' in 'ambulant' zu 3/75 Fehlern.

Aber auch mit vollständiger Affixerkennung dürfte die Fehlerquote für fremdstämmige Wörter immer noch wesentlich höher als für deutschstämmige liegen, denn Zerlegung und Transkription sind an den deutschstämmigen Wörtern ausgerichtet. Da aber deutschstämmige Wörter wesentlich häufiger auftreten als fremdstämmige, sollten sich, auf Texte bezogen, für beide Wortarten ungefähr gleiche Fehlerquoten ergeben.

Bei der Transkription von Fremdwörtern zeigten sich drei hauptsächlich Fehlerquellen in den Umsetzungsregeln:

- 'ie' wird auch in Fremdwörtern als langes 'i' realisiert, wenn es in 'IE' oder 'IŠ' umgesetzt werden sollte, z.B. in 'Studie', 'orientiert' und 'Klient'.
- 'V' wird zwar in deutschstämmigen Wörtern am Wortanfang meistens in 'F' umgesetzt, in Fremdwörtern dagegen immer in 'W'. Die fehlende Unterscheidung führte zu 9/50 Fehlern.
- 'AE', 'OE' und 'UE' codieren in deutschen Wörtern zwar die Umlaute, können in Fremdwörtern aber auch zwei Vokale darstellen.

6.2 Die Bestimmung des Wortakzents

Die Bestimmung des Akzentmusters für Wörter orientiert sich an den Arbeiten von Bierwisch (1966) und Kiparsky (1966) sowie an den im Duden (Mangold 1974a) und von Siebs (1969) angegebenen Betonungsregeln.

Von Bierwisch und Kiparsky habe ich das Konzept übernommen; ein Regelsystem zu entwickeln, das eine unbeschränkte Anzahl von Nebenakzenten zuläßt, so daß entsprechend dem in Kapitel 4 beschriebenen Modell deutscher Wörter beliebig komplizierte Wortstrukturen akzentuiert werden können. Dabei dürfen jedoch im Gegensatz zu den von Bierwisch und Kiparsky entwickelten Modellen Nebenakzente gleichen Grades auftreten. Die im Duden und bei Siebs angegebenen Betonungsregeln waren Grundlage für eine erste Fassung der Akzentregeln. Diese Regeln basieren auf zwei für Menschen leicht zu erfüllenden, aber für Maschinen u.U. problematischen Voraussetzungen:

- Der Anwender der Regeln muß feststellen können; ob ein Wort deutscher oder fremdstämmiger Herkunft ist; weil die Regeln zur Akzentuierung deutscher Wörter völlig anders sind als die fremdstämmiger Wörter.
- Er muß die Wortstruktur bestimmen können; d.h. er muß Stämme, Vorsilben, Nachsilben und Endungen erkennen können, um die Akzentsilben zu finden.

Das schon beschriebene Programm zur Analyse der Wortstruktur ist in der Lage, beide Punkte zu erfüllen, den ersten jedoch nur durch einen Kunstgriff. Da Fremdwörter in deutschem Text nur ca. 4% der Wörter ausmachen, ist für sie wegen der geringen Häufigkeit eine höhere Fehlerquote zulässig. Außerdem sind die meisten Fremdwörter endbetont, d.h. der Akzent liegt auf einer der letzten Silben des Wortes. Die Betonungsregeln für Fremdwörter wurden deshalb darauf reduziert, bestimmte akzenttragende wortfinale Silben zu identifizieren, die von der Affixerkennung, wie schon beschrieben, wie Nachsilben behandelt werden können. Durch die Verifizierung, die die Affixerkennung durchführt; ist sichergestellt, daß kaum Verwechslungen mit wortfinalen Silben von deutschstämmigen Wörtern auftreten.

Dieses Vorgehen unterscheidet sich grundsätzlich von dem im Duden und bei Siebs angegebenen, weil nicht mehr zwei Regelsätze für zwei verschiedene Wortfamilien verwendet werden, statt dessen gibt es nur einen Regelsatz, der in seinen Grundzügen den im Du-

den angegebenen Akzentregeln für deutschstämmige Wörter entspricht. Dieser Regelsatz wird ergänzt durch eine Identifizierung von akzenttragenden Wortteilen aus Fremdwörtern mit so hoher Sicherheit, daß praktisch eine Umkehrung des Lösungsweges gegenüber dem Ansatz der Wörterbücher vorliegt: Nicht, weil ein Wort fremdstämmig ist, erhält eine bestimmte Buchstabenfolge einen Akzent, sondern weil eine Buchstabenfolge als akzentuiert bestimmt worden ist, kann man schließen, daß ein Fremdwort vorliegt.

Zur Bestimmung des Wortakzents muß also die Wortstruktur bekannt sein, und Vor- und Nachsilben müssen identifiziert werden, damit über ihre Akzentuierbarkeit entschieden werden kann. Daraus ergibt sich eine Zweiteilung des Regelsatzes. Es werden Regeln gebraucht, die anhand der identifizierten Vor- und Nachsilben über die Akzentuierbarkeit der Affixe entscheiden, und weitere Regeln, die aus der vollständig bekannten Wortstruktur das Akzentmuster bestimmen.

6.2.1 Die Akzentuierbarkeit von Affixen

Der Duden führt drei Regeln zur Akzentuierung deutscher Vorsilben an:

- Die Präfixe be-, ent-, er-, ge-, ver- und zer- sind nicht betont.
- Die Partikeln ab-, an-, aus-, bei-, ein-, nach- und wieder- sind meistens betont.
- Die Partikeln da-, dar-, durch-, her-, hier-, hin-, hinter-, in-, miß-, ob-, über-, um-, un-, unter-, voll-, vor-, wider- und zu- kommen betont und unbetont vor.

Siebs stellt die Regeln ähnlich dar, zusätzlich werden zur zweiten und dritten Regel für jede Vorsilbe Listen der Wörter mit betonten, unbetonten und wechselnd betonten Vorsilben angegeben.

Beispiele

Akzentregel

betont

1. Die Vorsilben be-, ent-, er-, ge-, ver- und zer- sind nicht akzentuierbar.

unbetont

2. Andere Vorsilben werden immer akzentuiert; wenn eine der unter 1. genannten Vorsilben folgt.

Überheblichkeit

Durchfall

3. Die Vorsilben durch-, über- und unter- tragen in Wörtern mit den Nachsilben -ung, -heit und -keit keinen Akzent, sonst sind sie akzentuiert.

herkommen

4. Die Vorsilben her- und hin- sind vor -an, -ab; -aus; -ein, -über; -unter und -vor nicht akzentuiert, sonst erhalten sie einen Akzent.

aufeinander

5. Die Vorsilben ab-, an-, um-, auf-, aus-, bei-, vor- und nach- werden in Kombinationen mit -einander nicht akzentuiert; sonst wohl.

Unsinn

Gegenstand - gegenüber

6. (Überwiegend) akzentuiert sind un-, ur-, zu-, ein-; mit-, gegen-, aber-, miss- und wider-.

danach - damals

7. (Überwiegend) ohne Akzent sind da-, in-, ob-, so- und dar-.

Liste 6.1: Regeln zur Akzentuierung von Vorsilben

Nur die erste Regel ist wegen ihrer Eindeutigkeit für Vorleseautomaten implementierbar, nicht jedoch die mehrdeutigen anderen Regeln. Da der Umfang der in der Umsetzung verwendeten Listen klein gehalten werden sollte, wurde untersucht, ob es unter Zu-

lassung einer gewissen Fehlerquote möglich ist, aus diesen Listen Regeln zu extrahieren. Das Ergebnis dieser Untersuchung sind die Regeln 2 bis 7 in Liste 6.1. Bis auf die beiden letzten Regeln wird eine niedrige Fehlerquote erreicht. Die beiden letzten Regeln sind als Auffangregeln zu verstehen, so daß der größere Teil der von ihnen behandelten Vorsilben den richtigen Akzent erhält, einige Vorsilben aber noch falsch akzentuiert werden. Die beiden letztgenannten Regeln könnten verbessert werden, wenn es möglich wäre, Substantive, Adjektive, Verben und Adverben zu unterscheiden. Da es dafür noch keinen geeigneten Algorithmus gibt, wurde darauf verzichtet.

Während die Akzentuierbarkeit von Vorsilben hauptsächlich durch Kontext-Abhängigkeiten im Wort bestimmt wird, ist sie bei Nachsilben bis auf einige noch nicht behandelte Ausnahmen nur durch den Nachsilbentyp bestimmt. Die in Liste 6.2 angeführten Regeln geben deshalb auch nur an, welche Nachsilben akzentuierbar sind und welche nicht.

Bei den Nachsilben gibt es einige, die sowohl als fremdstämmige akzentuierbare Nachsilbe als auch als deutschstämmige nicht akzentuierbare Nachsilbe oder Endung auftreten können, z.B. '-em' und '-in':

Phonem - jedem
Leiterin - Vitaminin.

Fallen Nachsilben mit Endungen zusammen, werden sie immer als Endungen behandelt, Ausnahmen sind nur möglich, wenn aufgrund des grammatikalischen Aufbaus des Wortes eine Endung ausgeschlossen werden kann, also wenn z.B. hinter '-em' Flexionsendungen auftreten. Fallen deutsch- und fremdstämmige Nachsilben zusammen; werden die fremdstämmigen Nachsilben in häufig vorkommenden Wörtern mit Hilfe von Umfeldbetrachtungen identifiziert, akzentuiert und ggf. abweichend phonemisiert. In selten auftretenden Wörtern führen sie dagegen zu Fehlern.

Die gesamte Information über die Akzentuierbarkeit von Vor- und Nachsilben ist in den Regelsätzen abgelegt, welche die Affix-

Beispiele

Akzentregel

Partei

halbieren

1. Die deutschen Suffixe -ei und -ieren werden akzentuiert.

Möglichkeit

nahrhaft

2. Alle anderen deutschen Suffixe können keinen Wortakzent tragen.

praktikabel

banal

Kommissar

flexibel

Motiv

Fraktur

3. Die folgenden fremdstämmigen Suffixe werden akzentuiert: -abel; -aille, -ade; -age, -al; -an, -and, -ant, -anz, -ar, -at, -eau; -ell; -enz, -ese, -esk; -ess, -ett; -ibel; -id, -ie; -ier; -ik, -il; -ion; -it; -itis, -ismus, -ist; -iv; -thek; -oge, -ose, -ur

Liste 6.2: Regeln zur Akzentuierung von Nachsilben

erkennung zur Identifizierung von Affixen benötigt, und sie wird vom Interpreter dem erkannten Affix zugeordnet. Ein Akzent wird aufgrund dieser Informationen aber erst vom Programm zur Bestimmung des Akzentmusters unter Berücksichtigung weiterer Regeln gesetzt.

6.2.2 Die Festlegung des Akzentmusters

Nachdem die vollständige Wortstruktur bekannt ist, verwendet ein hierarchischer Satz von Regeln die Information über den Wortaufbau und über die Akzentuierbarkeit von Affixen, um ein Akzentmuster zu bestimmen.

Dazu werden zunächst alle Stämme mit einem Akzent auf dem Vokal der ersten Silbe versehen. Besteht das Wort nur aus einem Stamm, eventuellen Endungen und nicht akzentuierbaren Vor- und/oder Nachsilben, wird der Akzent in einen Hauptakzent umgewandelt. Weitere Akzente werden in solchen einfachen Wörtern nicht ge-

setzt. Mit diesen Regeln wird bereits der größte Teil der Wörter in laufendem Text korrekt akzentuiert, und zwar fast alle einfachen deutschstämmigen Wörter und Lehnwörter. Außerdem werden die Regeln auf Fremdwörter angewendet, die keine zu akzentuierende Nachsilbe enthalten. Dort führen die Regeln allerdings häufig zu Fehlern; denn Fremdwörter werden im Regelfall nicht auf der ersten Silbe des Stammes betont.

Werden in einem Teilwort des Wortmodells II in Kapitel 4 mehrere akzentuierbare Nachsilben gefunden, wird nur ein Hauptakzent auf die Letzte gesetzt. Außerdem wird bei einer akzentuierbaren Nachsilbe der Hauptakzent auf dem vorhergehenden Stamm um eine Stufe zu einem Nebenantwort erster Ordnung reduziert. Wird dann innerhalb des Teilwortes noch eine akzentuierbare Vorsilbe gefunden, erhält sie den Hauptakzent, und die bereits gesetzten Akzente werden um eine Stufe vermindert.

Wenn eine Wortzusammensetzung aus mehreren Teilwörtern besteht, werden zunächst alle Teilwörter nach den o.a. Regeln akzentuiert, danach werden die Akzente entsprechend der Stellung des Teilwortes in der Zusammensetzung vermindert, und zwar im ersten Teilwort gar nicht, im zweiten um eine Stufe, im dritten um zwei Stufen usw. Bei dieser Regel wurde außer acht gelassen, daß aufgrund der Zuordnung der Teilwörter zueinander auch andere Akzentmuster möglich sind, z.B. kann das Wort 'Dampfschiffahrt' je nach Akzentmuster verschiedene Bedeutungen haben:

Dampfschiffahrt = Schiffahrt mit Dampf
1 2 3

Dampfschiffahrt = Fahrt mit Dampfschiff
1 3 2

Trotzdem ist eine Verbesserung dieser Regel nicht vordringlich, denn die seltenen Fehler dieser Art fallen zwar beim Hören auf, sind aber nicht verständnismindernd.

In Liste 6.3 sind die Regeln noch einmal mit Beispielen zusammengefaßt.

1. Stämme erhalten einen Akzent auf dem Vokal der ersten Silbe.

Pause Ferien

2. Sind keine akzentuierbaren Vor- und/oder Nachsilben und nur ein Stamm vorhanden, wird nur ein Hauptakzent auf dem Stamm gesetzt.

Vergeßlichkeit beachten

3. Folgen mehrere akzentuierbare Nachsilben hintereinander, wird nur die letzte betont.

Bibliothek Bibliothekar

4. Sind akzentuierte Nachsilben vorhanden; erhalten sie den Hauptakzent, der Akzent auf dem Stamm wird um eine Stufe zu einem Nebenakzent reduziert.

agieren Litanei
2 1 2 1

5. Sind akzentuierbare Vorsilben vorhanden; erhalten sie den Hauptakzent, Akzente auf Stamm und Nachsilben werden um eine Stufe reduziert.

Nebenfach unakzentuiert
1 2 1 3 2

6. Zusammengesetzte Wörter tragen den Hauptakzent auf dem ersten Teilwort, im zweiten sind alle Akzente um eine Stufe vermindert, im dritten um zwei Stufen usw.

Dampfschiffahrt Leihbücherei
1 2 3 1 3 2

Liste 6.3: Regeln zur Bestimmung des Akzentmusters

6.2.3 Fehler bei der Bestimmung des Wortakzents

Im untersuchten Korpus (siehe Kap. 4.4) wurden 81/314 Akzentuierungsfehler gefunden, was einer Fehlerquote von 7,3/3,5 % entspricht. Die meisten dieser Fehler traten in Fremdwörtern auf und waren auf nicht in den Listen enthaltene, betonbare Nachsilben zurückzuführen. Außerdem sollte die Vorsilbenliste um fremdstämmige Vorsilben wie 'pro-', 'kon-', 'kom-', 'inter-' usw. ergänzt werden, damit auch Wörter wie z.B. 'Problem' richtig akzentuiert werden.

Deutsche Wörter werden größtenteils korrekt akzentuiert (23/85 Fehler), bis auf Wörter, die mit 'über-', 'unter-' und 'durch-' beginnen. Die dritte Regel in Liste 6.1 sollte deshalb verbessert werden. Außerdem sollte die Vorsilbenliste erweitert und auch Wörter wie 'voraus-' und 'zusammen-' aufgenommen werden.

15/108 falsch akzentuierte Wörter waren auch falsch phonemisiert, wobei für beide Fehler die gleiche Ursache, nämlich eine falsche oder unvollständige Affixabtrennung, vorlag. Unter Berücksichtigung dieser Doppelfehler ergeben sich insgesamt 120/488 falsch umgesetzte Wörter, das entspricht einer Fehlerquote von 10,8/5,5 %.

6.3 Zusammenfassung

Es wurde ein Transkriptionsverfahren vorgestellt, das in der Lage ist, sowohl deutsch- als auch fremdstämmige Wörter in Lautschrift umzusetzen. Aufgrund des Ablaufs der Transkription in mehreren Schritten und aufgrund der vorhergehenden Zerlegung von Wörtern in Morpheme kommt die Umsetzung mit nur ca. 70 Transkriptionsregeln aus und benötigt so um Größenordnungen weniger Regeln als eine Transkription auf Wortebene.

Die Bestimmung des Wortakzents läuft ebenfalls in mehreren Schritten ab. Zunächst werden Vor- und Nachsilben bestimmt und über deren Akzentuierbarkeit entschieden. Mit Hilfe der Angaben

über die Akzentuierbarkeit von Affixen errechnet ein nur aus wenigen Regeln bestehender Algorithmus Akzentmuster für einzelne Wörter. Wortzusammensetzungen werden akzentuiert, indem zunächst die Akzentmuster für die Teilwörter bestimmt werden. Diese Muster werden zusammengefaßt und aneinander angepaßt.

Die Regeln sind für den deutschstämmigen Wortschatz entwickelt worden, können aber auch fremdstämmige Wörter transkribieren. Da aber noch nicht alle fremdstämmigen Affixe erkannt werden, macht die Fehlerquote in fremdstämmigen Wörtern derzeit noch mehr als drei Viertel der Gesamtfehlerquote aus.

7. INTONATION UND SYNTHETISATORANSTEUERUNG

Der Kern der vorliegenden Arbeit wurde in den vorherigen Kapiteln beschrieben, nämlich die Entwicklung einer umfassenden Struktur zur Umwandlung von Text in gesprochene Sprache sowie im Besonderen die Entwicklung von Algorithmen zur Umwandlung von Text in Lautschrift und Akzent auf der Wortebene. Die Behandlung von Text auf Satzebene und die Erzeugung der Satzintonation, für die u.a. der Wortakzent benötigt wird, sind Teil einer Folgedissertation. Näheres darüber findet sich bei Kulas und Rühl (1982, 1984, ca.1984).

Eine der Randbedingungen der Arbeit war jedoch, ein System zu entwickeln, das auch während der Entwicklung schon praktisch einsetzbar sein sollte. Aus diesem Grund, und weil eine akustische Ausgabe des erzeugten Akzents sowieso sinnvoll ist, wurden Verfahren für eine einfache Intonation und Verfahren zur Ansteuerung von Phonemsynthetisatoren entwickelt. Die einfache Wortintonation wird im Rahmen der Folgedissertation durch eine Satzintonation ersetzt.

7.1 Ein einfaches Betonungsmodell

Während die bei der Wortumwandlung erzeugte Lautschrift direkt an einen Phonemsynthetisator übergeben werden kann und dann eine, wenn auch monotone Aussprache erzeugt, muß der Wortakzent erst in eine Wortintonation umgerechnet werden, ehe damit ein Synthetisator angesteuert werden kann.

Intonation hängt nicht nur vom Wortakzent, sondern noch von vielen anderen auf der Wortebene unbekannten Parametern ab. Deshalb wurde nur ein einfaches Verfahren zur Bestimmung von Grundfrequenz und Vokaldauer als Funktion des Wortakzents entwickelt, das zur Kontrolle des Wortakzents ausreicht, weil es näherungsweise die Intonation eines isoliert gesprochenen Einzelwortes erzeugt. Zur Erzeugung einer vollständigen Satzintonation ist es

aber ungeeignet.

Zur Bestimmung des Grundfrequenzverlaufs habe ich mich an den Arbeiten von O. von Essen (1964) und von Isacenko und Schädlich (1966) orientiert. Von Essen hat postuliert, daß deutsche Sätze einen allgemein abfallenden Grundfrequenzverlauf haben. Dieser abfallende Grundfrequenzverlauf wurde auf das Modell der isoliert gesprochenen Wörter übertragen.

Isacenko und Schädlich haben mit einem Zwei-Stufen-Modell des Grundfrequenzverlaufs in einem Satz experimentiert, und sie haben nachgewiesen, daß eine Grundfrequenzänderung direkt vor, in oder direkt nach einem Vokal von Hörern als Betonung auf diesem Vokal empfunden wird.

Abfallender Grundfrequenzverlauf und starke Frequenzänderung auf dem zu betonenden Vokal wurden in einem Verlauf der Grundfrequenz als abfallende Treppenfunktion zusammengefaßt; wobei die Übergänge zwischen den Stufen allerdings nicht abrupt sind, sondern sich über die Dauer des zu betonenden Vokals hinziehen. Zur Unterscheidung von Haupt- und Nebenakzenten ist dem Hauptakzent eine Frequenzänderung um zwei Halbtöne zugeordnet, dem ersten und zweiten Nebenakzent eine Änderung von nicht ganz einem Halbton. Alle weiteren Nebenakzente ändern die Grundfrequenz nicht, sondern nur die Lautdauer, falls sie überhaupt realisiert werden. Die Größen der Grundfrequenzänderungen wurden empirisch gewonnen. Aufgrund dieser Maßnahmen ergibt sich der in Bild 7.1 gezeigte schematische Grundfrequenzverlauf für Wörter.

Neben der Grundfrequenz wurde auch die Dauer der langen Vokale als Funktion des Akzents geändert, denn laut Duden sind die langen Vokale in unbetonter Stellung kürzer als in betonter Stellung. Lange Vokale in unbetonter Stellung oder mit der dritten oder einer schwächeren Nebenbetonung werden deshalb um ca. 40% gegenüber langen Vokalen mit Haupt- oder Nebenbetonung gekürzt. Auch dieser Wert wurde empirisch ermittelt.

Es hat sich gezeigt, daß diese Maßnahmen zur Erzeugung einer Wortbetonung für isoliert gesprochene Einzelwörter ausreichen,

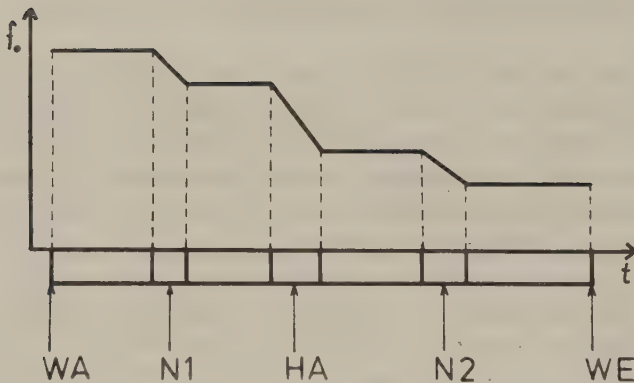


Bild 7.1: Schematischer Verlauf der Grundfrequenz als Funktion des Wortakzents. Es bedeuten:

- WA : Wortanfang
- WE : Wortende
- HA : Hauptakzent
- N1 : 1. Nebentakzent
- N2 : 2. Nebentakzent

aber die Verkettung von Einzelwörtern zu Sätzen noch sehr unnatürlich klingt. Verantwortlich sind dafür nicht so sehr die Tonbrüche, d.h. die abrupten Grundfrequenzänderungen zwischen den Wörtern, als vielmehr eine Überbetonung der Formwörter, die in fließender menschlicher Sprache üblicherweise unbetont sind und schneller ausgesprochen werden als Sinnwörter.

Deshalb wurden zwei weitere phänomenologische Maßnahmen getroffen, die eigentlich in den Bereich der Satzintonation gehören. So wurde die Akzentrealisierung bei Formwörtern unterdrückt; d.h. die Grundfrequenz ändert sich nicht und lange Vokale werden gekürzt. Danach werden alle Phoneme des Formwortes um 30% in ihrer Länge gekürzt, so daß z.B. für lange Vokale die gesamte Kürzung 60% ausmacht. Der Faktor der Kürzung wurde empirisch ermittelt. Diese Verbesserungen erzeugen zwar noch keine natürliche Aussprache, insbesondere keine Satzmelodie, sie führen aber zu einer für

eine Zwischenlösung akzeptablen Sprachqualität.

7.2 Die Schnittstelle zum Phonemsynthetisator

Die bei derzeit kommerziell erhältlichen Synthetisatoren vorhandenen Schnittstellen zur Datenübergabe sind noch sehr uneinheitlich, wobei insbesondere die Möglichkeiten zur Übergabe der intonatorischen Parameter Grundfrequenz, Lautdauer und Intensität sehr stark voneinander abweichen und meist unvollständig sind. So können z.B. bei den beiden am Lehrstuhl verwendeten Synthetisatoren Votrax VS6-G und Votrax SC01 der Fa. Federal Screw Works neben den Phonemen nur vier Grundfrequenzstufen übergeben werden, und es ist möglich, bei den Vokalphonemen zwischen bis zu drei Varianten mit verschiedenen Dauern auszuwählen. Die vier sehr grob gestuften Grundfrequenzen (Frequenzverhältnisse: 1,0 : 1,24 : 1,34 : 1,46) können zwar für eine Zwischenversion zur Realisierung einer Wortbetonung verwendet werden, sind aber für die Satzbetonung ungeeignet. Das Gleiche gilt für die bis zu drei verschiedenen Phonemdauern. Eine ausführliche Untersuchung der Schnittstellen von derzeit erhältlichen Synthetisatoren ist in (Rühl, Dreissig und Kulas 1984) durchgeführt worden. Wegen der Unzulänglichkeiten der bestehenden Schnittstellen und um nicht auf einen bestimmten Synthetisator festgelegt zu sein, wurde eine programminterne Schnittstelle definiert, die bei der Berechnung der Intonation keinerlei Beschränkung auferlegt, und die durch ein geeignetes Treiberprogramm an die meisten erhältlichen Synthetisatoren anpassbar ist.

7.2.1 Datenformate der Schnittstelle

Formant- und LPC-Synthetisatoren werden angesteuert; indem sie zu festen Zeitpunkten - üblicherweise alle 5 bis 25 ms - einen neuen Datensatz (engl.: "frame") erhalten. Analog dazu wurde als Bezeichnung für einen an der Schnittstelle zu übergebenden Datensatz der Begriff "Datenrahmen" gewählt. Anders als bei den o.a. Synthetisatoren hat ein Datenrahmen der Schnittstelle jedoch kei-

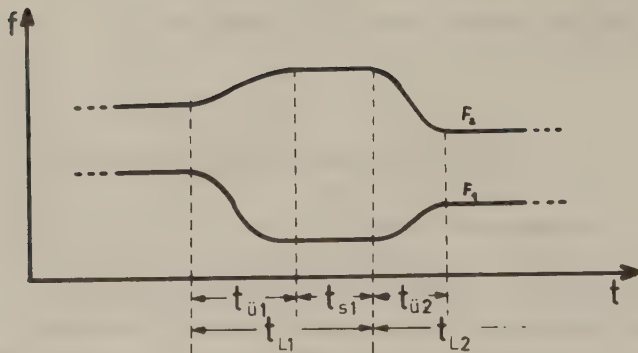


Bild 7.2: Zeitlicher Aufbau eines Lautes. Es bedeuten:

F_1 , F_2 : zeitlicher Verlauf der Formanten 1 und 2

t_{L1} , t_{L2} : Dauer der Laute L_1 , L_2

t_{s1} : Dauer des stationären Teils des Lautes L_1

$t_{Ü1}$, $t_{Ü2}$: Dauer des Übergangs der Laute L_1 , L_2

ne feste, sondern eine variable Dauer, die neben den Codierungen der Lautschrift, der Grundfrequenz und der Intensität Teil des Datenrahmens ist. Jeder Datenrahmen umfaßt vier Byte; wobei in jedem Byte ein Datenwert übergeben wird.

Der Lautwert der Phoneme wird in der programminternen Lautschrift entsprechend Anhang 6 codiert, die das Treiberprogramm auf den Lautvorrat des verwendeten Synthetisators abbildet. Ausgenommen davon sind die beiden durch ein Zeichen codierten Lautfolgen 'K/' und 'T/', die vor der Übergabe in 'K', 'S' bzw. 'T', 'S' umcodiert werden müssen. Da bisher nur ca. 50 verschiedene Laute vom Programm erzeugt werden, ist bei 256 möglichen Lautcodierungen noch genügend Platz für Erweiterungen vorhanden.

Bild 7.2 zeigt schematisch die Struktur eines Phonems im Zeitbereich. Sie wird durch das Verhalten der Formanten bestimmt. Ein Laut besteht üblicherweise aus einem stationären Teil, während dessen Dauer sich die Formanten nicht oder nur wenig ändern, und einem Übergang, in dem sich die Formanten vom stationären Teil des vorhergehenden Phonems auf den stationären Teil des folgenden Phonems hin bewegen. Ein stationärer Teil muß nicht vorhanden

sein, die Übergänge zwischen verschiedenen Phonemen dürfen sich sogar überlappen. Im Rahmen der Schnittstelle ist die Gesamtdauer eines Phonems entsprechend Bild 7.2 als die Zeit zwischen den Anfangszeitpunkten zweier Übergänge definiert. Die Dauer eines Phonems muß nicht gleich der Dauer eines Datenrahmens sein, sondern ein Phonem kann durch mehrere Rahmen dargestellt werden.

Für alle Laute außer der langen Pause wird die Rahmendauer in 2-ms-Schritten codiert, es sind also Lautdauern zwischen 2 und 510 ms einstellbar. Diese Maximaldauer und die Feinheit der Auflösung sollten für alle Laute außer der langen Pause ausreichen. Die lange Pause dagegen wird, wie der Name schon sagt, zur Darstellung von langen, semantisch und syntaktisch bedingten Pausen benötigt, die bei langsamer Sprechweise mehr als fünf Sekunden dauern können. Für diese Pausen wurde deshalb die grobere, aber ausreichende Auflösung im 32-ms-Raster gewählt; woraus sich eine Maximaldauer von 8,16 Sekunden ergibt.

Beim Sprechen reicht der Grundfrequenzbereich der Stimme von ca. 55 Hz bei Männern mit tiefer Stimmlage bis zu 300 Hz bei Kleinkindern. Einzelne Sprecher verwenden einen Grundfrequenzumfang von ca. 8 Halbtönen bei neutraler, ca. einer Oktave bei emotionaler Sprechweise und in Extremsituationen bis zu zwei Oktaven. Intonationserzeugende Programme sollten sich deshalb mit 8 Halbtönen bescheiden. Die Schnittstelle sollte aber mindestens einen Grundfrequenzumfang von 2 Oktaven erlauben, damit durch verschiedene Stimmlagen auch mehrere Sprecher simuliert werden können, wenn es der Synthetisator zuläßt.

Zur Skalierung der Grundfrequenz wurde die Frequenzmaßskala, also eine Tonhöhenintervallskala gewählt. Tonhöhen-skalen haben den Vorteil, daß sie an der Tonhöhenempfindung des Menschen ausgerichtet sind. Eine logarithmische Skala wurde genommen, weil auf ihr die Grundfrequenzänderungen in der Sprache besser dargestellt werden können. Grundfrequenzänderungen zur Erzeugung einer Satzmelodie oder als Folge von Akzenten lassen sich nämlich besser durch Frequenzverhältnisse als durch Frequenzdifferenzen beschreiben; denn wenn z.B. bei unterschiedlichen Grundfrequenzen ein Akzent realisiert werden soll, ändert sich die Frequenz-

differenz, nicht aber das Frequenzverhältnis.

Nach Isacenko und Schädlich (1966) reicht eine Grundfrequenzänderung von einem Halbton aus, um den Eindruck einer Betonung zu erzeugen, noch geringere Frequenzänderungen werden nicht mehr als Betonung empfunden. In menschlicher Sprache treten aber geringere Grundfrequenzänderungen auf, z.B. als 'intrinsic pitch', d.h. als lautspezifische Grundfrequenzänderung. Diese Frequenzänderungen liegen bei einigen Prozent (Neweklowsky 1975, Antoniadis und Strube 1981), machen die Sprache aber klarer und verständlicher. Um auch diese Änderungen noch sauber darstellen zu können, wurden 10 Cent als kleinste Skalierungseinheit gewählt, was einer Frequenzänderung von 0,58% entspricht. Mit dieser Einheit entsprechen 10 Stufen einem Halbton und 120 Stufen einer Oktave, so daß mit 255 Werten ein Bereich von mehr als zwei Oktaven darstellbar ist.

Durch den ausreichend großen Wertebereich kann die Grundfrequenz absolut übergeben werden. Wäre dies nicht möglich gewesen, hätte sie auch als Differenz zum vorhergehenden Frequenzwert übergeben werden können, denn zwischen zwei Datenrahmen ändert sich die Grundfrequenz höchstens um einige Halbtöne. Um den gesamten männlichen Stimmbereich überstreichen zu können, den die meisten derzeit erhältlichen Synthetisatoren unterstützen, wurde als Bezugstonhöhe der Ton 'A1' mit 55 Hz gewählt, der höchste übergebbare Ton mit dem Wert 255 ist dann 'ais' + 50 Cent.

Aus Gründen des Komforts hat der eigentlich den Grundton 'A1' darstellende Wert '0' eine Sonderbedeutung erhalten. Wird er übergeben, interpretiert ihn das Treiberprogramm dahingehend, daß die Grundfrequenz gegenüber dem vorhergehenden Laut nicht geändert wird.

Obwohl nach Isacenko und Schädlich stufige Änderungen der Grundfrequenz durchaus zulässig sind, zeigten am Lehrstuhl durchgeführte Hörversuche eine Minderung der Sprachqualität durch sprunghafte Grundfrequenzübergänge. Deshalb ist ein Grundfrequenzverlauf, wie er in Bild 7.3 dargestellt ist, nicht zulässig, zwischen den Stufen muß geeignet interpoliert werden. Leider ist

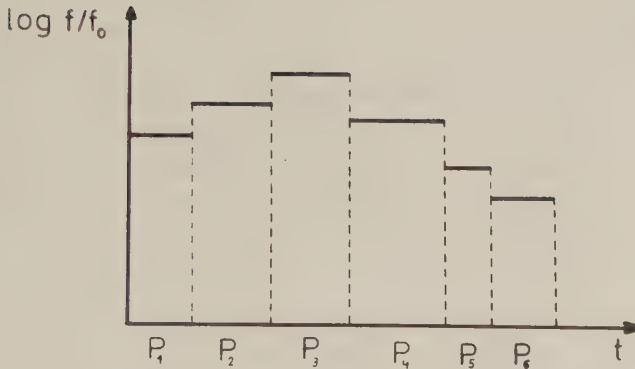


Bild 7.3: Grundfrequenzverlauf mit stufiger Änderung der Frequenz

P_1 bis P_6 : Datenrahmen 1 bis 6

dies nicht bei allen Phonemsynthetisatoren möglich, bei Synthetisatoren ohne Interpolationsmöglichkeit muß deshalb die Erzeugung des "intrinsic pitch" durch das Umsetzungsprogramm unterdrückt werden, um den Grundfrequenzverlauf möglichst kontinuierlich zu halten. Bei möglicher Interpolation wird die Grundfrequenz für das Ende des Datenrahmens angegeben, und während der Rahmendauer wird zwischen der Grundfrequenz des vorhergehenden Rahmens und der angegebenen Frequenz interpoliert. Die Interpolation erfolgt, wie in Bild 7.4 dargestellt; entlang einer Geraden.

Für die Intensität reicht eine Abstufung in 1-dB-Schritten völlig aus, denn dies ist in etwa die Intensitätsdifferenz, die menschliche Hörer gerade noch feststellen können. Da die Intensität als Funktion der Betonung nur wenig schwankt; wird nur ein Dynamikbereich von ca. 20 dB benötigt. Die Intensität wird als "Dämpfung" übergeben, so daß '0' die höchste und '20' die niedrigste Intensität codiert. Als ziemlich unkritische Größe ist sie stufig für die Dauer des gesamten Rahmens definiert. Es ist jedoch darauf zu achten, daß die Intensität durch einen Tiefpaß an geeigneter Stelle im Phonemsynthetisator "weich" geändert wird, damit kein Umschaltknacken zu hören ist. Ein Tiefpaß erster Ordnung mit einer Zeitkonstante zwischen 20 ms und 40 ms ist dafür

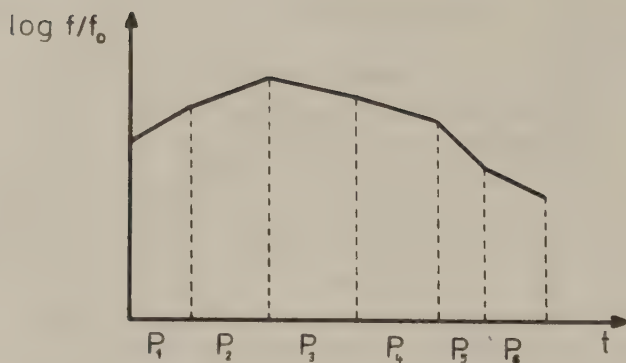


Bild 7.4: Grundfrequenzverlauf mit interpolierter Änderung der Frequenz

P_1 bis P_6 : Datenrahmen 1 bis 6

geeignet.

7.2.2 Der Zusammenhang zwischen Phonemen und Datenrahmen

Wenn nicht aufgrund von Akzenten eine besondere Bedingung vorliegt, haben Grundfrequenz und Intensität einen ziemlich kontinuierlichen Verlauf; so daß es möglich ist, die meisten Laute durch einen Datenrahmen darzustellen. Als Folge von Akzenten treten jedoch stärkere Grundfrequenzänderungen auf, die sich über mehrere Phoneme hinziehen können, und bei denen die Krümmung der Grundfrequenzkurve so stark ist, daß die Annäherung durch eine Gerade pro Phonem zu Qualitätsminderungen führt. In solche Fällen ist es möglich, Laute durch mehrere Rahmen darzustellen. So könnten z.B. die Rahmen P_2 und P_3 in Bild 7.4 zu einem betonten Vokal mit einem Grundfrequenzmaximum gehören, und der Rahmen P_1 zu einem vorhergehenden Konsonanten, während dessen Verlauf die Grundfrequenz bereits anfängt zu steigen. Um Parameterverläufe exakter nachzubilden, ist es theoretisch möglich, die Anzahl der Rahmen pro Laut beliebig zu steigern.

Praktisch ist die Anzahl der Datenrahmen durch die Teile des

Zwischenspeichers für die Rahmen und durch den Aufwand begrenzt, den man für das Treiberprogramm in Kauf nehmen will, denn je nach verwendetem Phonemsynthetisator müssen unter Umständen alle Datenrahmen für einen Laut abgespeichert werden, ehe sie in Ansteuerdaten umgesetzt werden können. Für die in diesem Kapitel beschriebene Wortbetonung reicht ein Rahmen pro Phonem aus. Doch für die Satzintonation, die derzeit am Lehrstuhl im Rahmen einer Folgedissertation entwickelt wird, wurde eine Zerlegung von Phonemen in bis zu vier Rahmen vorgesehen.

7.2.3 Das Interface des Phonemsynthetisators VS6-G

Wie die vorhergegangene Betrachtung der Ansteuermöglichkeiten für Phonemsynthetisatoren zeigt, ist es nicht ohne weiteres möglich, die in der internen Schnittstelle übergebenen Daten an kommerziell erhältliche Synthetisatoren anzupassen. Es kann aber in vielen Teilen das Hardware-Interface des Synthetisators so modifiziert werden, daß eine genügend feine intonatorische Steuerung erreicht wird. So wurden Schaltungsmodifikationen für den Votrax SC01 und den Votrax VS6-G und entsprechende Treiberprogramme entwickelt. Diese sollen am Beispiel des VS6-G näher erläutert werden.

Eine ausführliche Beschreibung des VS6-G findet sich bei Soskuty (1977). Ohne Modifikationen können pro Laut nur bis zu drei unterschiedliche Dauern eingestellt werden, die üblich im Verhältnis 1 : 1,6 : 2,6 abgestuft sind, und es können vier Grundfrequenzen in der Abstufung 1 : 1,24 : 1,34 : 1,46 gewählt werden. Auf der Frontplatte sind aber Drehknöpfe angebracht, die über Potentiometer eine manuelle Steuerung aller prosodischen Parameter in weiten Bereichen erlauben. Zum Schutz ihres Know-Hows hat die Fa. Votrax jedoch alle Platinen im VS6-G vergossen, und die Dokumentation des technischen Aufbaus ist im Handbuch so knapp gehalten, daß kein direkter Eingriff in die Schaltung möglich schien.

Um die Intonation trotzdem fein steuern zu können, wurde deshalb erwogen, die Potentiometer auf der Frontplatte durch digital

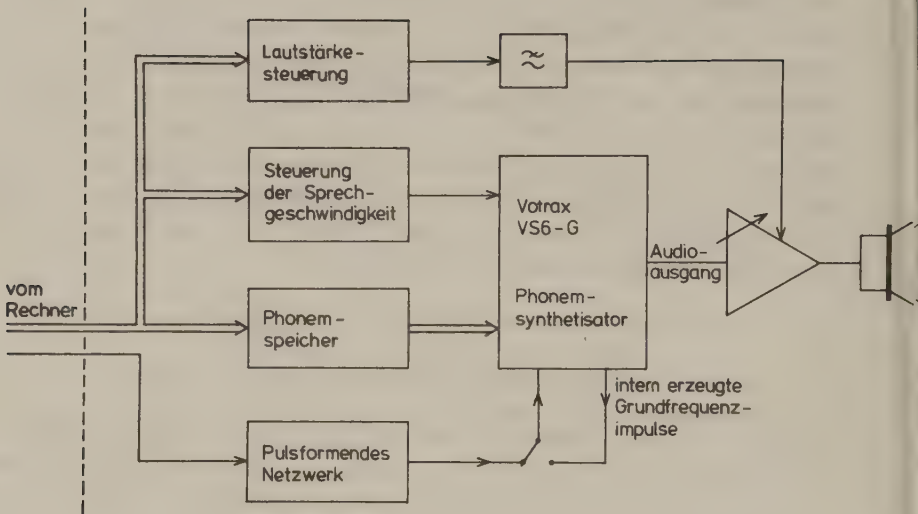


Bild 7.5: Blockschaltbild des modifizierten Votrax VS6-G

steuerbare elektronische Potentiometer zu ersetzen. Diese Lösung wurde für Intensität und Sprechgeschwindigkeit ausgeführt, aber für die stimmhafte Anregung wurde eine Leitung gefunden; auf der die Anregungspulse von der stimmhaften Quelle zur Vokaltrakt-Simulation übertragen werden. Diese Leitung wurde aufgetrennt, und die Anregung wurde durch einen neuen Pulsgenerator mit gleicher Pulsform ersetzt, bei dem jeder Puls vom Treiber getriggert werden muß. Dadurch erhöht sich zwar die Rechenbelastung des steuernden Rechners, aber der Abstand der Anregungsimpulse wird fast im Mikrosekundenraaster aufgelöst, und die maximale Grundfrequenz wird nur durch die Schnelligkeit des Treibers begrenzt.

Da bei der bisher verwendeten Parallelschnittstelle mit Zwischenspeicher für 80 Phoneme unsicher war; wann welches Phonem ausgegeben wurde, und dadurch keine direkte Zuordnung der Intonationsparameter zu den Phonemen möglich war; wurde eine eigene Schittstellenkarte mit einem nur ein Byte tiefen Speicher entwickelt, auf der auch die Schaltungen zur Intonationssteuerung untergebracht wurden, und die in einem noch freien Steckplatz des VS6-G Platz findet. Bild 7.5 zeigt das Blockschaltbild des derart

modifizierten Gerätes. Eine ausführliche Beschreibung der Schaltungen findet sich bei Dreissig, Kulas und Rühl (1984).

Durch die Schaltungsmodifikationen wird die Ausgabe von Lautschrift, Lautdauer und Intensität sehr einfach, sie können bis auf einige einfache Zusatzregeln für Diphthonge, spezielle Durcodierungen und ähnliches über Tabellen in Ansteuerdaten für den Synthetisator umgewandelt werden. Die Grundfrequenzsteuerung dagegen benötigt einigen Rechenaufwand, weil bei Frequenzänderungen entlang einer Geraden im Cent/Zeit-Bereich interpoliert wird, und diese Interpolation für jeden Impuls durchgeführt werden muß.

Lineare Interpolation der Grundfrequenz im Cent-Zeit-Bereich bedeutet exponentiellen Anstieg oder Abfall der Grundfrequenz:

$$F(t) = F_0 \exp(at) \quad (I)$$

Daraus ergibt sich die Periodendauer mit $T_p = 1/F$ zu

$$T_p(t) = T_0 \exp(-at) \quad (II)$$

(I) und (II) sind widersprüchliche und nicht realisierbare Beschreibungen des Interpolationsvorgangs, denn die diskrete Periodendauer T_p ist zugleich das Zeitintervall T der kontinuierlichen Zeit t und damit zu verschiedenen Zeiten innerhalb des Intervalls verschieden groß. Man erhält aber eine realisierbare Funktion, wenn man auf eine diskrete Darstellung übergeht und die Periodendauer T_p für das Zeitintervall T konstant hält. Die Diskretisierung ist zugleich eine Annäherung an die dem Modell zugrundeliegenden physikalischen Vorgänge, denn die Glottis als anregende Quelle für den menschlichen Vokaltrakt sendet tatsächlich eine Folge von Impulsen aus, deren zeitliche Abfolge durch (II) aufgrund der kontinuierlichen Darstellung nur näherungsweise beschrieben wird. Die gesuchte Funktion ist also eine Folge von Impulsen konstanter Amplitude, aber mit variablem Abstand, wobei sich der Abstand zweier Impulse zu den Zeiten t_n und t_{n+1} als die Periodendauer $T(t_n)$ ergibt; die man bei Einsetzen von t_n in (II) erhält.

$$t_{n+1} = t_n + T(t_n) \quad (\text{III})$$

$T(t_n)$ berechnet sich in rekursiver Schreibung aus Gleichung (III):

$$\begin{aligned} T(t_n) &= T_0 \exp(-at_n) \\ &= T_0 \exp(-a(t_{n-1} + T(t_{n-1}))) \\ &= T(t_{n-1}) \exp(-aT(t_{n-1})) \end{aligned} \quad (\text{IV})$$

$$T(t_n) \approx T(t_{n-1}) - aT^2(t_{n-1}) \quad (\text{V})$$

Damit die Näherung durch (V) gültig ist; muß $aT(t_{n-1})$ klein gegen eins sein. Eine Worst-Case-Abschätzung zeigt; daß selbst für maximale Periodendauer ($f = 55 \text{ Hz}$) und die enorme Anstiegsgeschwindigkeit von 5 Okt/s die Größe aT kleiner als 0,1 und üblich kleiner als 0,01 ist. Die Interpolation ist damit auf zwei Multiplikationen und eine Addition zurückgeführt.

Durch diese Maßnahmen kann die Berechnung eines Impulsabstandes in weniger als einer Millisekunde durchgeführt werden. Den Ablauf einer Interpolation zeigt Bild 7.6. "Unter-" bzw. "Überschwinger" werden unterdrückt; und es braucht nur so lange interpoliert zu werden, wie die Zielperiodendauer noch nicht erreicht ist. Die Zielperiodendauer und der Koeffizient a müssen einmal pro Datenrahmen berechnet werden. Ändert sich die Grundfrequenz gegenüber dem vorhergehenden Rahmen nicht, werden die alten Werte übernommen. Durch die Übernahme auch des Koeffizienten a wird erreicht, daß eine wegen der Näherung noch nicht vollständig durchgeführte Interpolation im folgenden Datenrahmen mit der richtigen Steigung zum Zielwert kommt.

Die Konstante a muß pro Datenrahmen einmal berechnet werden; dazu gilt:

$$\frac{a}{(\text{s}^{-1})} = \frac{\Delta f}{(\text{Okt})} \cdot \ln 2 \cdot \frac{1/\Delta t}{(\text{s}^{-1})}$$

wobei Δf die Frequenzänderung im Datenrahmen und Δt die Länge des

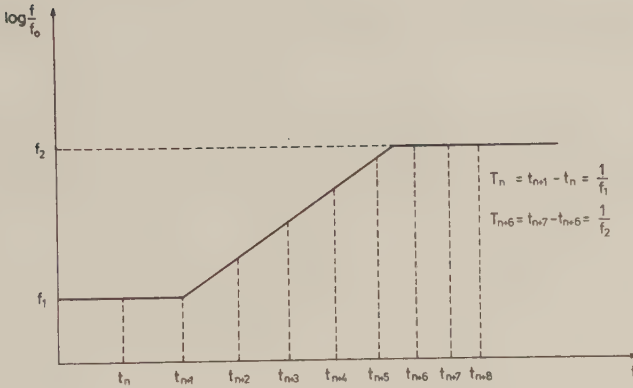


Bild 7.6: Interpolation der Grundfrequenz als Folge eines Grundfrequenzsprunges

Rahmens ist. Intern ist die Grundfrequenz in 10-Cent-Schritten und die Zeit in 2-ms-Intervallen codiert; und die Dauer der Periode T wird als eine Anzahl von Rechnertakten angegeben. Mit der Dauer eines Rechnertaktes $\tau = 1/0,9216 \text{ MHz}$ erhält man die zuge-schnittene Größengleichung:

$$\frac{a}{(\tau^{-1})} = 3,13 \cdot 10^{-6} \cdot \frac{\Delta f}{(10\text{ct})} \cdot \frac{1/\Delta t}{(2\text{ms}^{-1})}$$

7.3 Zusammenfassung

Es wurde ein vorläufiges Verfahren zur Umsetzung des Wortakzents in Wortintonation vorgestellt, das in Zwischenversionen des Vorleseautomaten den Wortakzent hörbar macht. Es erzeugt einen fallenden Verlauf der Grundfrequenz im Wort; wobei Grundfrequenzänderungen nur während der Dauer der akzentuierten Vokale stattfinden. Zusammen mit Maßnahmen zur Kürzung der Dauer von unbetonten langen Vokalen, der unterdrückten Akzentuierung für Formwörter und der Kürzung der Dauer von Formwörtern ist die Wortintonation in vorläufigen Versionen des Vorleseautomaten

nicht nur zur Betonung von Einzelwörtern, sondern auch zum Vorlesen von zusammenhängenden Texten geeignet.

Um zu verhindern, daß bei der Entwicklung von Verfahren zur Bestimmung der Intonation Abhängigkeiten vom verwendeten Phonemsynthetisator entstehen, wurde eine programminterne Schnittstelle definiert, die die Umsetzung von dem synthetisatorabhängigen Treiberprogramm trennt. Am Beispiel des Votrax VS6-G wurde der Anschluß eines Phonemsynthetisators an die Schnittstelle beschrieben.

8. DIE TEXTVORVERARBEITUNG

Text besteht nicht nur aus Buchstaben, sondern auch aus vielen anderen Zeichen wie Ziffern; Sonderzeichen wie '%' und '\$', Interpunktionen usw. Die Aussprache der Zeichen liegt oft nicht eindeutig fest, sondern ist vom Darstellungsformat abhängig (z.B. kann ein '-' ein Minuszeichen oder ein Worttrennzeichen sein). Verschiedene Formate können zur gleichen Aussprache führen, z.B. sollte die Aussprache unabhängig davon sein; ob Wörter nur mit Groß- oder Kleinbuchstaben geschrieben werden; wie oft rechnerintern üblich, oder ob sie orthografisch richtig geschrieben werden mit großen Anfangsbuchstaben für Substantive und am Satzanfang, sonst aber klein. Dagegen sind die Umsetzungsprogramme nur für ein Eingabeformat ausgelegt: nämlich großgeschriebene Wörter, die durch Zwischenräume getrennt werden. Um die praktisch unbegrenzte Vielfalt der Eingabeformate an die Umsetzung anzupassen, wird eine Textvorverarbeitung benötigt.

Wegen der vielfältigen Eingabeformate kann es keine vollständige Textvorverarbeitung geben, denn zur Bestimmung der Bedeutung von Zeichenfolgen und der daraus folgenden Aussprache reichen oft syntaktische und semantische Analysen nicht aus; und es muß auf pragmatisches Wissen zurückgegriffen werden (z.B. beim Vorlesen von Rechnerprogrammen). Deshalb wurde nur das Grundgerüst einer Textvorverarbeitung entwickelt, mit dem üblicher Text in Prosaform vorgelesen werden kann, spezielle aufgabenspezifische Formate werden nicht behandelt. Es wurde jedoch darauf geachtet; daß Routinen zur Umsetzung aller Grundformate vorhanden sind, so daß aufgabenspezifische Anpassungen und höhere Formate aus Grundformaten zusammengesetzt werden können. War das zugrundeliegende Format nicht eindeutig erkennbar, wurde eine Umwandlung gewählt, die möglichst eindeutig auf das vorliegende Textformat schließen läßt.

Bild 8.1 zeigt den Aufbau der Textvorverarbeitung: Sie besteht im wesentlichen aus einem Eingabeprogramm, das zeilenweise beliebigen Text einliest und innerhalb der Zeile die üblichen Editiermöglichkeiten bietet, der darauf folgenden eigentlichen Vor-

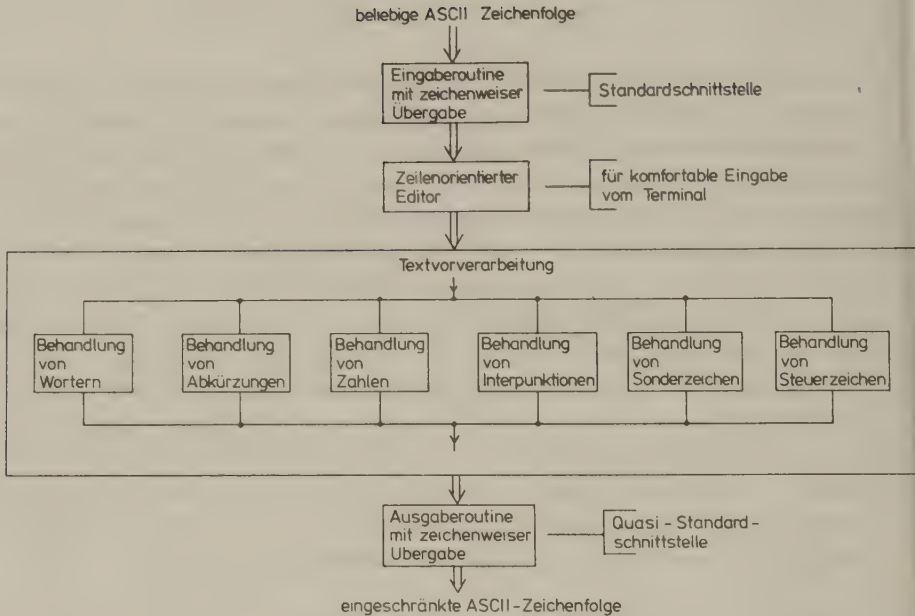


Bild 8.1: Aufbau von Texteingabe und -vorverarbeitung

verarbeitung, die den Text in ein Standardformat umwandelt, und der Ausgaberoutine, die den formatierten Text an die Umsetzungsprogramme weiterleitet. Die eigentliche Textvorverarbeitung besteht aus einer Reihe unabhängiger Programmteile, von denen jedes für die Formatierung einer bestimmten Klasse von Zeichen zuständig ist.

8.1 Die Behandlung von Schrifttext

Deutsche Texte werden üblicherweise mit Kleinbuchstaben geschrieben, nur am Satzanfang und am Anfang von Substantiven wird ein Großbuchstabe verwendet. Dies ist aber nicht immer der Fall. Z.B. wird im Telex-Verkehr nicht zwischen Groß- und Kleinschreibung unterschieden, und im Dialog mit Rechnern wird meistens nur

Groß- oder nur Kleinschreibung verwendet. Um alle möglichen Schreibweisen zu berücksichtigen; dürfen Wörter deshalb in beliebiger Form an die Sprachausgabe übergeben werden. So sind z.B. 'Wort', 'WORT', 'wort' und 'wOrT' gleichermaßen gültige Schreibweisen des Wortes 'Wort!'. Die Schreibweise 'Wort', d.h. die normale deutsche Rechtschreibung mit großen Anfangsbuchstaben für Substantive sollte jedoch bevorzugt verwendet werden, da diese Schreibweise zusätzliche Information über die Wortart enthält, die von der Textvorverarbeitung weitergeleitet werden und zu einer besseren Satzbetonung führen.

Abkürzungen werden derzeit nur in der Schreibung mit Abkürzungspunkten nach jedem Buchstaben erkannt und dann buchstabiert. Das heißt, Abkürzungen der 'F.D.P.' werden erkannt und als 'EFF DE PE' ausgesprochen, dagegen werden 'SPD', 'CDU' und 'Prof.' fehlerhaft behandelt. Auch isoliert stehende Buchstaben werden buchstabiert, d.h. 'm' wird als 'EMM' ausgesprochen und nicht als 'mmm'-Laut. Wörter und Abkürzungen dürfen sowohl die 26 Buchstaben des Alphabets als auch die deutschen Umlaute und das 'ß' enthalten.

Im deutschen Sprachraum werden in der Datenverarbeitung zwei unterschiedliche Zeichensätze verwendet: der amerikanische Standard des ASCII-Code und ein ASCII-Code mit deutschen Sonderbelegungen, im folgenden als ASCII/G bezeichnet. Die beiden Codes sind weitgehend identisch, das wesentliche Unterscheidungsmerkmal sind die Codierungen der Umlaute und das 'ß'; die im amerikanischen ASCII mit Sonderzeichen belegt sind.

Code (hex)	ASCII	ASCII/G
5B	[	Ä
5C	\	Ö
5D	]	Ü
7B	{	ä
7C		ö
7D	}	ü
7E	~	ß

Um den Vorleseautomaten möglichst universell einsetzbar zu machen, wurden Maßnahmen getroffen, damit beide Zeichensätze verarbeitet werden können. Es wurde ein per Software bedienbarer Umschalter vorgesehen, der 3 Stellungen annehmen kann: deutscher Zeichensatz, amerikanischer Zeichensatz und unbekannter Zeichensatz. Bei festgelegtem Zeichensatz werden die Zeichen entweder als Buchstaben behandelt oder als Sonderzeichen ausgesprochen. Ist der Zeichensatz dagegen unbekannt, stellt die Vorverarbeitung selbst anhand von Regeln fest, welcher Zeichensatz wahrscheinlich vorliegt. Dazu wird davon ausgegangen, daß Texte entweder in ASCII oder in ASCII/G geschrieben sind, d.h. die betroffenen Sonderzeichen treten nicht mit wechselnder Bedeutung auf. Deshalb wird an einem einmal gewählten Zeichensatz festgehalten und nur dann umgeschaltet, wenn hinreichend sicher gewährleistet ist, daß der bisher gewählte Zeichensatz falsch ist.

Als Gründe, vom deutschen Zeichensatz auf ASCII umzuschalten, werden dabei akzeptiert:

- Eines der Zeichen tritt alleinstehend auf, d.h. das Wort hat keine Buchstaben als Nachbarn und würde nur aus diesem Zeichen bestehen.

- Es tritt ein 'ß' am Wortanfang auf.

Akzeptable Gründe zur Umschaltung von ASCII auf ASCII/G sind:

- Eines der Zeichen tritt in einem Wort auf; d.h. rechter und linker Nachbar sind Buchstaben.

- Eines der Zeichen außer '[' und '{' tritt am Wortanfang auf.

- Eines der Zeichen außer ']' und '}' tritt am Wortende auf.

Diese Regeln sind in der Lage, schnell eine Änderung des Zeichensatzes zu erkennen, können jedoch nicht verhindern, daß nach einem Wechsel des Zeichensatzes manchmal Fehler auftreten. So wird z.B. das Wort 'Ärger' als "ECKIGE KLAMMER AUF RGER" ausgesprochen, wenn von ASCII in den deutschen Zeichensatz gewechselt wird.

selt wurde und dieser Wechsel vom Programm noch nicht erkannt ist.

8.2 Die Behandlung von Zahlen

Von der Zahlenumsetzung werden vier unterschiedliche Zahlenformate unterschieden.

1. ganze Zahlen
2. Ordnungszahlen
3. Dezimalzahlen
4. Uhrzeiten

Ganze Zahlen können wie '123456' zusammengeschrieben sein oder aber wie '123 456' und '123.456' durch Punkte bzw. Leerzeichen optisch in Dreierblöcke gegliedert sein. Andere Formatierungen wie die bei Telefonnummern übliche Gliederung in Zweierblöcke (z.B. 17 24 64) sind nicht berücksichtigt, die Zahl wird dann in drei einzeln ausgesprochene Zahlengruppen aufgebrochen (siebzehn vierundzwanzig vierundsechzig). Unberücksichtigt ist auch eine wechselnde und nicht übliche optische Gliederung durch Punkte und Leerzeichen.

Ein führendes '-' vor einer Zahl wird als 'Minus' ausgesprochen, ein führendes '+' als 'Plus'. Führende Nullen werden nicht unterdrückt, sondern ausgesprochen. Da führende Nullen bei Zahlen in normalem Text unüblich sind, liegt wahrscheinlich ein Sonderformat vor, in dem die Nullen bedeutungstragend sind; z.B. die Vorwahl einer Telefonnummer.

Die Zahlen zwischen 1100 und 1999 werden nicht als Tausender-Zahlen, sondern als Hunderter-Zahlen ausgesprochen (z.B. 1981 = NEUNZEHNHUNDERTEINUNDACHTZIG), weil wahrscheinlich eine Jahreszahl vorliegt.

Die größte aussprechbare Zahl ist die Zahl 999 999 999 999. Größere Zahlen werden "buchstabiert"; wenn sie nicht optisch ge-

gliedert sind. Bei optisch gegliederten Zahlen werden die Tausenderblöcke so lange einzeln ausgesprochen, bis ein von der Länge her aussprechbares Zahlenformat vorliegt. Die einzelnen Tausenderblöcke werden als ganze Zahlen ausgesprochen, wenn die Zahl durch Leerzeichen gegliedert ist, bei Gliederung durch Punkte werden sie als Ordnungszahlen ausgesprochen.

Beispiele:

9990000000000	neun neun neun null null null null null null null null null null
999 100 000 000 000	neunhundertneunundneunzig einhundert Milliarden
999.100.000.000.000	neunhundertneunundneunzigste einhundert Milliarden

Ordnungszahlen werden wie ganze Zahlen behandelt und an die ganze Zahl dann die Endung 'ste' bzw. 'te' gehängt. Ausgenommen sind die Zahlen '1', '3' und auf '01' bzw. '03' endende Zahlen. Hier wird 'eins' bzw. 'drei' ersetzt durch 'Erste' bzw. 'Dritte'. Verboten (und unüblich) ist die optische Gliederung von Ordnungszahlen durch Punkte. In diesem Fall wird jede durch einen Punkt abgeschlossene Zahlengruppe als Ordnungszahl ausgesprochen.

Bei Dezimalzahlen wird der Vorkomma-Teil wie eine ganze Zahl behandelt und ausgesprochen. Außerdem ist es zulässig, daß der Vorkomma-Teil durch einen Bindestrich ersetzt wird (z.B. -,79). In diesem Falle wird für den Bindestrich NULL ausgesprochen. Auch der Fall, daß der Nachkomma-Teil der Zahl durch einen Bindestrich ersetzt wurde, ist zulässig, Komma und Bindestrich werden dann überlesen und die Zahl als ganze Zahl ausgesprochen.

Besteht der Nachkomma-Teil aus Ziffern, gibt es zwei mögliche Formen der Aussprache: Wenn der Nachkomma-Teil nur zwei Ziffern lang ist, wird er als Zehnerblock ausgesprochen (z.B. 0,79 = Null Komma Neunundsiebzig) in den anderen Fällen wird jede Ziffer einzeln ausgesprochen (z.B. 0,3190 = Null Komma Drei Eins Neun Null).

Während die o.a. Formate ähnlich behandelt werden; handelt es sich bei der Uhrzeit um ein Sonderformat. Damit eine Uhrzeit als solche erkannt wird, muß sie aus einer oder zwei Ziffern, einem Punkt und weiteren zwei Ziffern bestehen. An diese Ziffernfolge muß dann, durch kein bis vier Leerzeichen getrennt, der allein stehende große oder kleine Buchstabe 'H' oder das Wort 'Uhr' anschließen. Diese Zeichenfolge wird dann als (Zahl) 'Uhr' (Zahl) ausgesprochen.

In keinem der o.a. Zahlenformate darf ein Zeilenende auftreten, auch im Uhrzeit-Format nicht. Sollte das Wort 'Uhr' des Uhrzeit-Formates am Anfang einer neuen Zeile stehen; wird die Uhrzeit ausgesprochen als (Ordnungszahl) (Zahl) 'Uhr'.

Andere Formate wie z.B. Datumsangaben werden ebenfalls von den beschriebenen Formaten durch Aufspaltung in kleinere Gruppen korrekt behandelt (z.B. 1.12.1981 = Erste Zwölfte Neunzehnhunderteinundachtzig) oder aber zumindest auf eine eindeutige Weise wiedergegeben (z.B. Dezimalklassifikation von Kapiteln: 1.3.2. = Erste Dritte Zweite, oder Brüche: $1/8$ = Eins Schrägstrich Acht).

8.3 Die Behandlung von Sonderzeichen und Interpunktionen

Als Sonderzeichen bzw. Interpunktionen werden die ASCII-Zeichen mit der hexadezimalen Codierung von 20 bis 2F, 39 bis 40, 50, 5E und 5F, sowie im amerikanischen ASCII von 5B bis 5D und 76 bis 7E verstanden. Die meisten dieser Zeichen sind eindeutige Sonderzeichen, die von der Textvorverarbeitung zur Aussprache durch einen String ersetzt werden (z.B. '\$' durch 'DOLLAR') Eine Aufstellung dieser Zeichen findet sich in Anhang 7.1.

Das Leerzeichen ' ' wird als Worttrennungs-Zeichen behandelt und weitergeleitet; wenn es nach einem Buchstaben folgt. Wurde vorher schon ein Zeichen weitergeleitet, das ebenfalls als Wortende-Markierung behandelt wird, wird das folgende Leerzeichen unterdrückt (Zusammenfassung von mehreren Leerzeichen zu einem, wichtig z.B. für Tabellen).

Die Interpunktionen '!', '.', ';;', ':', ';', und '?' werden ebenfalls als Worttrennungs-Zeichen behandelt, folgen jedoch mehrere Interpunktionen aufeinander; was in normalem Text unüblich ist, wird nur die erste als Wortende und Satzteil- oder Satzende markierendes Zeichen aufgefaßt, die folgenden werden ausgesprochen.

Die beiden Anführungszeichen Apostroph "'" und Doppelapostroph "" werden unterdrückt. Nur wenn ein Apostroph innerhalb eines Wortes auftritt, kann es für die Transkription von Bedeutung sein und wird deshalb weitergeleitet.

Das Sonderzeichen '-', der Bindestrich, kann mit vielfacher Bedeutung in deutschen Texten auftreten. Er kann als sogenannter Gedankenstrich zwischen zwei Leerzeichen stehend - wie in diesem Beispiel - vorkommen; in diesem Fall wird er wie eine Interpunktion behandelt. Er kann in komplexen Wort-Zusammensetzungen zwischen Buchstaben stehend auftreten, in diesem Fall wird die Kennzeichnung einer Morphemgrenze weitergeleitet. Er kann am Wortanfang und am Wortende auftreten (Beispiele: Wortanfang und -ende, Wort- und Satzanfang), um Auslassungen zu kennzeichnen; in diesem Fall trägt er keine relevante Information für die Aussprache und wird unterdrückt.

Außer im Text kann der Bindestrich auch in Zahlen vorkommen. Steht er vor oder hinter einem Komma (z.B. -,79 oder 79,-); ersetzt er den fehlenden Vor- oder Nachkomma-Teil einer Zahl. Im ersten Fall wird er als 'NULL KOMMA ...' ausgesprochen; im zweiten Fall wird er unterdrückt. Steht der Bindestrich vor der ersten Ziffer einer Zahl; wird er als Vorzeichen aufgefaßt und als 'MINUS' ausgesprochen. Siehe dazu auch den Abschnitt über Zahlenumsetzung.

8.4 Die Behandlung von Steuerzeichen

Als Steuerzeichen werden die ASCII-Zeichen von 00hex bis 1Fhex verstanden, weil sie nicht visuell dargestellt werden, sondern zur Steuerung von Datenfluß und Eingabe- und Ausgabegeräten dienen. Die meisten Steuerzeichen werden von der Textvorverarbeitung unterdrückt, es werden nur einige Zeichen zur Textformatierung, zur Steuerung der Texteingabe und zur Steuerung der Sprachausgabe ausgewertet.

Die Untersuchung des Textformats beschränkt sich auf die Bestimmung des Zeilenendes; das durch Wagenrücklauf 'CR' oder Zeilenvorschub 'LF' als auch durch die Folge 'CR; LF' gekennzeichnet sein kann. Die Berücksichtigung aller üblichen Formen des Zeilenendes ist wichtig, damit einerseits keine zusätzlichen Leerzeilen erzeugt werden, die zu Aussprachefehlern bei Worttrennungen führen können, andererseits aber die Kennzeichnung von Absätzen durch Leerzeilen nicht unterdrückt wird, weil sie für die Intonation wichtige Information enthält.

Einige Steuerzeichen werden für die Texteingabe ausgewertet, um das letzte eingegebene Zeichen oder die letzte eingegebene Zeile zu löschen.

Einige weitere Steuerzeichen werden für die Sprachausgabe benötigt, um die letzte eingegebene Zeile erneut aussprechen zu lassen, um die Sprachausgabe per Software neu zu initialisieren und um verschiedene Parameter wie z.B. Lautstärke, Stimmlage und Sprechgeschwindigkeit ändern zu können.

Eine Aufstellung der verwendeten Steuerzeichen findet sich in Anhang 7.2.

8.5 Zusammenfassung

Die Textvorverarbeitung filtert den von der Eingabeschnittstelle kommenden Zeichenstrom und setzt alle für die Sprachausgabe nicht verarbeitbaren Zeichen und Zeichenfolgen in nicht verarbeitbaren Formaten in 'aussprechbare' Zeichenfolgen um:

- Wörter werden in Großschreibung umgesetzt.
- Abkürzungen werden buchstabiert.
- Zahlen; die in verschiedenen Formaten als ganze Zahl, Dezimalzahl, Ordnungszahl; Jahreszahl; Datum oder Uhrzeit eingegeben werden können; werden in eine ausgeschriebene Form umgewandelt.
- Sonderzeichen werden ausgeschrieben.
- Über Escape-Sequenzen werden Parameter für die Sprachausgabe geändert.

Die Textvorverarbeitung kann alle Zeichenformate bearbeiten; die in 'normalem' Text; also z.B. in Zeitungsartikeln auftreten. Mit Hilfe der dabei verwendeten Formatierungsbausteine läßt sie sich einfach an aufgabenspezifische Formate anpassen.

9. ERGEBNISSE

9.1 Überblick über die geleisteten Arbeiten

Aus Untersuchungen über das Vorlesen beim Menschen wurde das Modell eines Vorleseautomaten abgeleitet. Für das Modell wurden zunächst verschiedene Verfahren zur Umsetzung von Wörtern kritisch verglichen, dann wurde ein verfeinertes Konzept der Einzelwortumsetzung entwickelt; das eine grammatikalische Analyse von Wörtern und morphembezogene Verfahren zur Umwandlung von Text in Lautschrift und zur Erzeugung des Wortakzents enthält. Bild 9.1 zeigt eine Zusammenfassung des Modells und des Konzepts der Wortumsetzung.

Die grammatikalische Analyse basiert auf einem von W. Kästner entwickelten Verfahren, mit dessen Hilfe die Vokallänge als Funktion des folgenden Konsonantenclusters bestimmt werden kann. Das Kästnersche Verfahren der Clusteranalyse wurde derart modifiziert und erweitert, daß es nun mit ausreichender hoher Wahrscheinlichkeit Morphemgrenzen in Wortzusammensetzungen feststellt. Da das Verfahren auf Morphemgrenzen zwischen Stämmen optimiert ist, mußte es durch Algorithmen zur Erkennung und Klassifizierung von Vorsilben, Nachsilben und Endungen ergänzt werden. Vorsilben und Nachsilben werden über Listen erkannt, die neben dem Affix selbst, seiner Lautschrift und einer Betonungsinformation auch noch Information enthalten; mit deren Hilfe Affixe von zu Stämmen gehörigen gleichen Zeichenketten unterschieden werden. Ergänzt wird die Analyse von einer Ausnahmenerkennung über Listen. Die Listen enthalten derzeit nur für die Satzbetonung wichtige Formwörter; sollen aber auch falsch transkribierte und falsch akzentuierte Wörter aufnehmen.

Wegen der vorhergehenden grammatikalischen Analyse von Wörtern konnten für die Umsetzung von Text in Lautschrift und zur Erzeugung des Wortakzents sehr einfache Verfahren entwickelt werden, die auf morphembezogenen Regeln basieren und sich an die Regelsätze anlehnen, die in sprachwissenschaftlichen Werken wie z.B.

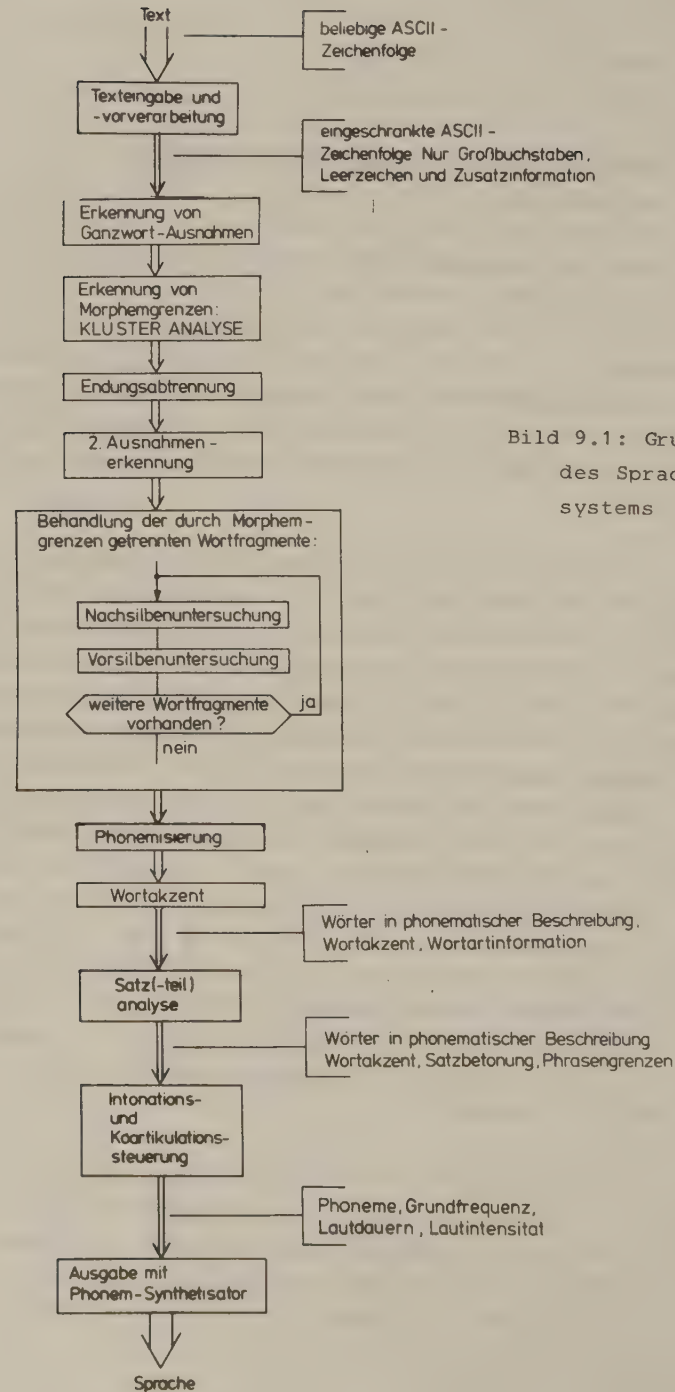


Bild 9.1: Grundstruktur
des Sprachsynthese-
systems

dem Duden angegeben werden.

Die Programme zur Wortumsetzung werden ergänzt durch Textvorverarbeitung und Ausgabeprogramme. Die Textvorverarbeitung bildet den ASCII-Zeichenvorrat in beliebigem Format auf den eingeschränkten Zeichensatz mit festem Format ab, den die Umsetzungsprogramme verarbeiten können. Für die Ausgabe wurde eine Schnittstelle mit Lautschrift und Intonationsparametern definiert, die die Entwicklung von Verfahren zur Intonationserzeugung vom verwendeten Phonemsynthetisator unabhängig macht. Dazu wurde beispielhaft das Treiberprogramm für einen Synthetisator vorgestellt. Ein einfaches Programm zur Wortintonation sowie vorläufige Maßnahmen zur Satzintonation schließen die Umsetzungsprogramme an die Schnittstelle an und vervollständigen damit eine Version des Vorleseautomaten.

9.2 Realisierungen von Vorleseautomaten

Der beschriebene Vorleseautomat existiert bisher in zwei Hardware-Ausführungen. Das Labor- und Entwicklungssystem besteht aus mehreren teilweise gekauften; teilweise selbst entwickelten Platinen und kann verschiedene Phonemsynthetisatoren ansteuern. Derzeit sind ein Votrax VS6-G und ein Votrax SC01 angeschlossen und ansteuerbar, aber zu Versuchszwecken waren auch schon andere Synthetisatoren angeschlossen. Das System arbeitet mit einem Mikroprozessor vom Typ M 6800. Hierauf laufen Programmversionen für mehrere Synthetisatoren und mit unterschiedlichen Betonungsalgorithmen, u.a. der beschriebenen Wortbetonung.

Die zweite Ausführung besteht aus der in Bild 9.2 gezeigten einfachen Europakarte. Sie enthält einen Mikroprozessor M 6803, serielle und parallele Eingabeschnittstellen; 16 KByte fast vollständig belegten Speicherplatz im EPROM; 4 KByte RAM-Speicherplatz und als Synthetisator einen Votrax SC01. Als Betonungsalgorithmus wird bei diesem Vorleseautomaten nicht mehr die einfache Wortbetonung verwendet; sondern eine auf die Wortbetonung aufbauende Satzmelodiesteuerung. Das durch den einfachen und kom-

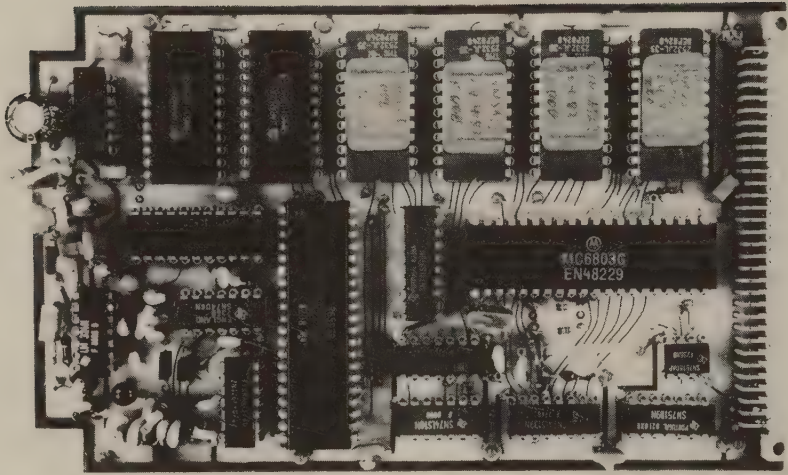


Bild 9.2: Realisierung des Vorleseautomaten auf einer Einfach-Europakarte mit dem Votrax SC01 als Phonemsynthetisator.

pakten Aufbau und die Verwendung eines IC's als Synthetisator erreichte gute Preis-Leistungs-Verhältnis war Anlaß für zwei Firmen, Adaptionen dieser Karte in ihren Blindenhilfsmitteln zu verwenden.

9.3 Ausblick

Der augenblickliche Entwicklungsstand des Vorleseautomaten wird dadurch charakterisiert, daß alle Algorithmen entwickelt worden sind, die auf der Ebene einer Umsetzung ohne Satzbetonung nach dem vorgestellten Konzept benötigt werden. Es wurde nachgewiesen, daß die Algorithmen ihre Aufgabe im Rahmen des Konzepts erfüllen, und weiterhin, daß das gewählte Konzept zu brauchbaren Realisierungen von Vorleseautomaten führt. Damit sind die Arbeiten zur grundsätzlichen Entwicklung der Verfahren abgeschlossen.

Auch aus der Sicht der Benutzer des Vorleseautomaten ist mit den dokumentierten Arbeiten ein Einschnitt erreicht. Die Qualität, insbesondere die Verständlichkeit von derzeit verfügbaren Phonemsynthetisatoren reicht noch nicht für Anwendungen von Vorleseautomaten mit ungeübten Benutzern aus; so daß die Verständlichkeit des Vorleseautomaten im Wesentlichen durch den Synthetisator bestimmt wird; weitere Verbesserungen der Umsetzungsprogramme auf Wortebene führen nur zu geringen Anhebungen der Verständlichkeit.

Daß die Qualität von Umsetzungsprogramm und Phonemsynthetisator derzeit in einem ausgewogenen Verhältnis stehen, darf aber nicht zu der Annahme verführen, daß keine Verbesserungen mehr notwendig sind. Im Hinblick auf zu erwartende Phonemsynthetisatoren mit besserer Sprachqualität sollte die Fehlerquote; d.h. die Anzahl der falsch akzentuierten bzw. falsch transkribierten Wörter gesenkt werden. Einen Hinweis, welche Fehlerquoten langfristig anzustreben sind, gibt die von der Fa. Digital Equipment gestellte Forderung für Vorleseautomaten, daß die 20 000 häufigsten Wörter fehlerfrei ausgesprochen werden sollten.

Für die vorgestellten Verfahren zur Wortumsetzung bedeutet das, daß zunächst die in dieser Arbeit erwähnten Verbesserungen durchgeführt werden. Dazu sind keine wesentlichen Änderungen der Algorithmen mehr notwendig, sondern nur noch Vervollständigungen von Datensätzen sowie Ergänzungen und Optimierungen von Regeln. Hieran wird zur Zeit weitergearbeitet.

Mit einem optimierten Programm ist zwar eine Fehlerquote deutlich unter fünf Prozent erreichbar; aber keine Fehlerfreiheit der häufigsten zwanzigtausend Wörter. Diese Tatsache scheint auf den ersten Blick das gewählte Konzept zu widerlegen. Doch ist zu bedenken, daß die o.a. Optimierung noch keine Erweiterungen der Ausnahmenlisten durch "echte", also zu Aussprache Fehlern führende Ausnahmen enthält. Gerade durch das gewählte Konzept mit grammatikalischer Zerlegung von Wörtern können Ausnahmen nicht nur in Form von ganzen Wörtern, sondern auch als Stamm, d.h. als Wort ohne Konjugations- und Deklinationseendung abgespeichert werden, und es ist möglich, noch weitere Zerlegungsebenen zu definieren,

auf denen Ausnahmenlisten angelegt werden können.

Die englische Sprache kennt im Gegensatz zur deutschen nur sehr wenige Konjugations- und Deklinationsendungen, so daß 20 000 fehlerfrei ausgesprochenen englischen Wörtern oder Wortformen ca. 60 000 bis 100 000 deutsche Wortformen entsprechen. Steht nur die Wortebene zur Abspeicherung von Ausnahmen zur Verfügung, müßten bei einer angenommenen Fehlerquote von 4 % ca. 4000 Ausnahmen erfaßt werden, um Fehlerfreiheit der häufigsten 100 000 Wörter zu erreichen. Bei durchschnittlich 20 Byte pro Eintrag würde dies zu einer Ausnahmenliste von 80 kByte Umfang führen. Trennt man jedoch die Endungen ab und führt damit Wörter auf gemeinsame Grundformen zurück, wird der benötigte Speicherplatz auf ca. ein Viertel reduziert und kann zusammen mit den Programmen im normalen Adressbereich eines 8-Bit-Prozessors untergebracht werden.

Durch die Abspeicherung von fehlerhaft ausgesprochenen Ausnahmen in Form von Wortteilen wird nicht nur eine Datenreduktion erreicht, sondern die Fehlerquote wird auch stärker gesenkt, als es mit nur einer Liste möglich wäre. Mit einer Liste, die ganze Wörter als Ausnahmen enthält, können nur die in der Liste enthaltenen Wörter korrigiert werden. Listen mit Wortteilen können dagegen auch Fehler in Wörtern korrigieren, die nicht zur Optimierungsmenge gehören, wenn sich die Wörter auf in den Listen enthaltene Grundformen zurückführen lassen.

Da die Umwandlung von Text in Ansteuerparameter für Phonemsynthesatoren mit einigen Prozent der Zeit auskommt, in der der Text ausgesprochen wird, sind nicht nur für die Satzbetonung, sondern auch für die Erweiterung der Listen keinerlei Zeitprobleme zu erwarten.

In Kapitel 2 wurden einige Anforderungen zusammengestellt, denen ein Vorleseautomat genügen sollte. Die Forderung nach maximaler Verständlichkeit und Natürlichkeit, der auf Wortebene eine möglichst kleine Fehlerquote bei der Umsetzung entspricht, kann durch die oben angeführten Verbesserungsvorschläge beliebig gut erfüllt werden. Echtzeitbetrieb ist durch die schnelle Umsetzung nicht nur im Monolog, sondern auch im Dialog durchführbar, und es

sind genügend Zeitreserven für Programmteile zur Satzbetonung vorhanden. Durch die Wortbetonung sind auch die derzeit lauffähigen vorläufigen Versionen nicht monoton und wirken nicht einschläfernd. Die bekannten Schnittstellen aus dem menschlichen Sprechverhalten wurden beachtet. Damit erfüllen die in dieser Arbeit vorgestellten Teile des Vorleseautomaten die in Kapitel 2 gestellten Anforderungen.

Literaturverzeichnis

- Allen, J.: Speech synthesis from unrestricted text. in: Flanagan, J.L.; Rabiner, L.R. (Eds.): Speech synthesis. Dowden, Hutchinson and Ross Inc, Stroudsburg (1973), 416-428
- Allen, J.; Hunnicutt, M.S.; Klatt, D.H.: MITalk-79. Notes for M.I.T. summer course 6.69s (1979)
- Allen, J.: Linguistic-based algorithms offer practical text-to-speech systems. Speech technology 1 (1981), 1, 12-16
- Antoniadis, Z.; Strube, H.W.: Untersuchungen zum "intrinsic pitch" deutscher Vokale. Phonetica 38 (1981), 277-290
- Atal, B.S.; Hanauer, S.L.: Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. 50 (1971), 637-655
- Berry-Rogghe, G.L.M.: Ein Programm zur automatischen Phonemisierung deutscher Texte. Unveröffentlichtes Manuskript, Mannheim (1976)
- Bierwisch, M.: Regeln für die Intonation deutscher Sätze. Studia Grammatica 7 (1966); 99-201
- Blauert, J.; Kulas, W.; Rühl, H.-W.; Schaffert, R.: Akustische Kommunikation an Arbeitsplätzen. Wirtschaftsverlag NW, Bremerhaven; in Vorbereitung (voraussichtl. 1984)
- Breuer, G. et al.: PHONTEXT - eine PHONOL-Anwendung zur Erzeugung synthetischer deutscher Sprache. Sprache und Datenverarbeitung 1 (1979), 2, 10-17
- Brustkern, J.; Willee, G.: Sprachsynthese mittels eines Computerprogramms: Das System Phontext. 4. DFG-Kolloquium Sprachverarbeitung, Darmstadt (1979); 83-84

- Carlson, R.; Granström, B.: A text-to-speech system based entirely on rules. Proc. ICASSP, Philadelphia (1976), 686-689
- Clark, H.H.; Clark, E.V.: Psychology and language. Harcourt Brace Jovanovich Inc., New York (1977)
- Class, F. et al.: Sprach-Ein-/Ausgabe. BMFT-Forschungsbericht, AEG-Telefunken, Ulm (1981)
- Coker, C.H.: Synthesis by rule from articulatory parameters. Proc. 1967 Conf. Speech Commun. Process. (1967), 52- 53
- Coker, C.H.: Speech synthesis with a parametric articulatory model. Speech Symposium, Kyoto (1968), paper A-4
- Coker, C.H.; Umeda, N.; Browman, C.P.: Automatic synthesis from ordinary English text. IEEE Trans. Audio. Elac. AU 21 (1973), 293-298
- Delmonte, R.: Automatic word-stress pattern assignment by rules: A computer program for standard Italian. Proc. 4th F.A.S.E. Symposium, Venedig (1981 a); 153-156
- Delmonte, R.: An automatic unrestricted text-to-speech prosodic translator. Proc. AICA, Padua (1981 b); 1075-1083
- Delmonte, R.: An automatic text-to-speech prosodic translator for Italian. in: Fortschritte der Akustik FASE/DAGA '82, Göttingen (1982); 1021-1026
- Deutsche Bundesbahn: Schlußbericht zum Forschungsvorhaben Fahrplanauskunft - Frankfurt/Main. München (1980)
- Dreissig, D.; Rühl, H.-W.; Kulas, W.: Sprachsynthese verfeinert. Elektronik (1984), H. 6, 68-70
- Dudley, H.: Synthesizing speech. Bell Laboratories Record 15 (1936), 98-102

- Dudley, H.; Riesz, R.R.; Watkins, S.A.: A synthetic speaker. J. Franklin Inst., (1939), 227, 739-764
- Dudley, H.: Remaking speech. J. Acoust. Soc. Am. 11 (1939), 169-177
- Dudley, H.; Tarnoczy, T.H.: The speaking machine of Wolfgang von Kempelen. J. Acoust. Soc. Am. 22 (1950), 151-166
- Elovitz, H.S. et al.: Letter-to-sound rules for automatic translation of English text to phonetics. IEEE Trans. ASSP-24 (1976), 6, 446-459
- Erben, J.: Deutsche Grammatik. Max Hueber Verlag, München (1980)
- Essen, O. von: Grundzüge der deutschen Satzintonation. A.Henn-Verlag, Ratingen/Düsseldorf (1964)
- Flanagan, J.L.: Note on the design of "terminal analog" speech synthesizers. J. Acoust. Soc. Am. 29 (1957), 306-310
- Flanagan, J.L.: Speech analysis, synthesis and perception. Springer, New York (1965)
- Flanagan, J.L.: Voices of men and machines. J. Acoust. Soc. Am. 51 (1972), 1375-1387
- Fromkin, V.: The non-anomalous nature of anomalous utterances. Language 47 (1971); 27-52
- Fromkin, V. (Ed.): Speech errors as linguistic evidence., Mouton, The Hague (1973)
- Garrett, M.F. : The analysis of sentence production. in Brower, G.H. (Ed.): The psychology of learning and motivation 9. Academic Press, New York (1975), 133-177
- Groner, G.F. et al.: A real-time text-to-speech converter. Speech technology 1 (1982), 2, 73-76

- Grossmann, E.: Mikroprozessor-gesteuerte Transkription von deutschem Text in Lautschrift. 4. DFG-Kolloquium Sprachverarbeitung, Darmstadt (1979), 88-89
- Grossmann, E. et al.: Synthese des unbegrenzten Vokabulars aus Text. in: Fortschritte der Akustik, DAGA '80, München (1980), 681-684
- Grossmann, E.: Speech synthesis of an unlimited vocabulary in the time domain from text. Proc. 4th F.A.S.E. Symposium on Acoustics and Speech, Venedig (1981); 157
- Grossmann, E.: Speech synthesis in the time domain from text. Proc. ICASSP 82, Paris (1982); 2; 936-939
- Harris, C.M.: A study of the building blocks in speech. J. Acoust. Soc. Am. 25 (1953), 962-969
- Holmes, J.N.; Mattingly, I.G.; Shearme, J.N.: Speech synthesis by rule. Language and Speech 7 (1964), 127-143
- Isacenko, A.V.; Schädlich, H.-J.: Untersuchungen über die deutsche Satzintonation. Studia Grammatica 7 (1966), 7-67
- Itakura, F.; Saito, S.: A statistikal method for estimation of speech spectral density and formant frequencies. Electron. Commun. Japan 53-A (1970); 36-43
- Itakura, F.; Saito, S.: On the optimum quantisation of feature parameters in the PARCOR speech synthesizer. Proc. 1972 Conf. Speech Commun. Process. (1972), 434-437
- Kästner, W.: Automatische Phonemisierung orthographischer Texte im Deutschen. Helmut Buske Verlag, Hamburg (1972)
- Kiparski, P.: Über den deutschen Akzent. Studia Grammatica 7 (1966); 69-98

- Klatt, D.H.: Software for a cascade/parallel formant synthesizer.
J. Acoust. Soc. Am. 67 (1980), 971-995
- Klatt, D.H.: The Klattalk text-to-speech conversion system. Proc.
ICASSP 82, Paris (1982), 3, 1589-1592
- Kommenda, M.: Entwicklung eines Sprachsynthesystems zur akustischen Ausgabe von deutschem Text. Diplom-Arbeit, Inst. f. Nachrichtentechnik d. TU Wien, Wien (1983)
- Kulas, W.; Rühl, H.-W.: SYNTEX - Synthetische Sprache aus Text. Eine Übersicht. NTG-Sitzung über Sprachsynthese, Berlin (1982a)
- Kulas, W.; Rühl, H.-W.: Satzzerlegung für ein Sprachsynthesystem mit unbegrenztem Wortschatz. in: Fortschritte der Akustik FASE/DAGA '82, Göttingen (1982b), 1017-1019
- Kulas, W.; Rühl, H.-W.: Über die Erzeugung einer Satzintonation. in: B.S. Müller (Ed.): Sprachsynthese, Georg Olms Verlag, im Druck (ca. 1984a)
- Kulas, W.; Rühl, H.-W.: Eine einfache Satzintonation für ein Sprachausgabesystem mit unbegrenztem Wortschatz. in: Fortschritte der Akustik DAGA '84, Darmstadt (1984b)
- Lewandowski, Th.: Linguistisches Wörterbuch. Quelle & Meyer, Heidelberg (1980)
- Lieberman, A.M. et al.: Minimal rules for synthesizing speech. J. Acoust. Soc. Am. 31 (1959); 1490-1499
- Mackensen, L.: Deutsches Wörterbuch. Pfahl-Verlag, Baden-Baden (1962)
- Mangold, M.: Der Laut. in: Duden - Die Grammatik. Duden-Verlag im BI, Mannheim (1973), 23-59

- Mangold, M.: Hochlautung. in: Duden - Aussprachewörterbuch, Dudenverlag im BI, Mannheim (1974 a); 29-61
- Mangold, M.: Deutsche Aussprachelehre. in: Duden - Aussprachewörterbuch, Dudenverlag im BI, Mannheim (1974 b), 69-104
- Mangold, H.; Stall, D.S.: Principles of text-controlled speech-synthesis with special application to German. in: Bolc, L. (Ed.): Speech communication with computers. Carl Hanser Verlag, München (1978), 139-181
- H. Mangold, Stall, D.S.; Zelinski, R.: Sprach-Ein-/Ausgabe. BMFT-Forschungsbericht; AEG-Telefunken; Ulm (1978)
- McIlroy, M.D.: Synthetic English speech by rule. Bell Tele. Lab. Inc., Murray Hill (1974)
- Meyer, H.: Deutsche Sprachstatistik. Georg Olms Verlagsbuchhandlung, Hildesheim (1964)
- Müller, B.S.: Regelgesteuerte Umsetzung von deutschen Texten in gesprochene Sprache für das Sprachausgabegerät VOTRAX. Selbstverlag GMD, St. Augustin (1981)
- Neweklowsky, G.: Spezifische Dauer und spezifische Tonhöhe der Vokale. *Phonetica* 32 (1975), 38-60
- Petersen, G.E.; Wang, W.S.-Y.; Sivertsen, E.: Segmentation techniques in speech synthesis. *J. Acoust. Soc. Am.* 30 (1958), 739-742
- Rühl, H.-W.: Ein Programm zur Phonemisierung orthographischer Texte mit Hilfe eines Mikroprozessors. Studienarbeit, Lehrst. f. allg. Elektrotechnik u. Akustik d. Ruhr-Univ. Bochum, Bochum (1977)

- Rühl, H.-W.: Ein Verfahren zur Zerlegung deutscher Wörter für die Sprachsynthese. 4. DFG-Kolloquium Sprachverarbeitung, Darmstadt (1979), 113-116
- Rühl, H.-W.: Über ein Wortanalyse-Verfahren für Vorleseautomaten. ACUSTICA 48 (1981a), 143-148
- Rühl, H.-W.: Wortbetonung für ein System zur Sprachsynthese aus Text. in: Fortschritte der Akustik DAGA '81, Berlin (1981b), 669-672
- Rühl, H.-W.; Blauert, J.: "S'NTEX" - the text-to-speech system of the Ruhr-University Bochum. A status report. Proc. 4th FASE Symposium on Acoustics and Speech, Venedig (1981c), 159
- Rühl, H.-W.: Ein Verfahren zur Bestimmung der Wortbetonung zur Sprachsynthese. ACUSTICA 51 (1982a), 12-17
- Rühl, H.-W.: SYNTEX - a microprocessor based system for automatic conversion of German text to speech. Proc. ICASSP 82, Paris (1982b), 1608-1611
- Rühl, H.-W.; Dreissig, D.; Kulas, W.: Sprachausgabe: Die Ansteuerung von Phonemsynthetisatoren. In Vorbereitung (vorraussichtl. 1984)
- Santos, J.M.; Nombela, J.R.: Text-to-speech conversion in Spanish. A complete rule-based synthesis system. Proc. ICASSP 82, Paris (1982), 3, 1593-1596
- Siebs, Th.: Deutsche Aussprache. De Gruyter, Berlin (1969), 42-143
- Slaby, W.A.: Entwicklung eines Programmsystems zur automatischen Umsetzung von Texten des Deutschen in gesprochene Sprache. BMFT-Forschungsbericht; Univ. Münster; Münster (1978)

- Soskuty, O.V.: Ein Sprachgenerator nach dem Prinzip der Phonem-synthese. Elektronik (1977), H. 9, 44-48
- Stall, D.S.: Morphologische Analyse deutscher Wörter zur automatischen Ermittlung der Lautschrift. 1. DFG-Kolloquium Sprachverarbeitung, Geisenheim (1974), 55-56
- Stall, D.S.: Ermittlung der Lautschrift deutschstämmiger Wörter für die Sprachsynthese nach Regeln. 2. DFG-Kolloquium Sprachverarbeitung, Vielbrunn (1976), 47-49
- Stall, D.S.: Zerlegung deutschstämmiger Wörter zur Ermittlung ihrer Aussprache. 3. DFG-Kolloquium Sprachverarbeitung, Muggendorf (1977); 70-71
- Synte: Synte 2. Datenblatt der Fa. Euroka Oy, Helsinki (1980)
- Uhle, M.: Die akustische Ausgabe orthografischer Texte. in: Fortschritte der Akustik DAGA '78, Bochum (1978), 473-476
- Votrax: Votrax audio response system; operator's manual. Handbuch zum VS6-G der Fa. Federal Screw Works, Troy (Michigan) (ca. 1976)
- Wolf, H.E.: Entwurf und Realisierung eines Formantsynthetisators mit paralleler Filterstruktur für die Sprachsynthese nach Regeln. Dissertation, Darmstadt (1981)
- Workshop am Heinrich-Hertz-Institut für Nachrichtentechnik Berlin: Aktuelle Fragen der Sprachsignalverarbeitung. Berlin (1979)
- Zimmermann, R.: Benutzerfreundliche Dialoggestaltung an Bildschirmarbeitsplätzen. NTZ 36 (1983), 5; 318-323

Anhang 1: Glossar der verwendeten linguistischen Begriffe

Bei der Erstellung des folgenden Glossars wurde für viele Angaben das "Linguistische Wörterbuch" von Th. Lewandowski (1980) zur Überprüfung der Richtigkeit verwendet.

Affix: ein -> Präfix, ein -> Suffix oder eine -> Flexionsendung; siehe auch -> Morphem.

Allophon: Realisierungsvariante eines -> Phonems. Die 'K'-Laute in "kalt" und "Kind" werden z.B. durch unterschiedliche Allophone realisiert; die sich in der Lage des Artikulationsortes, d.h. in der Position des Verschlusses unterscheiden.

Artikulation: Bildung von Lauten mit Hilfe des Sprechapparats; Einstellung; Haltung und Bewegung der Sprechwerkzeuge

Artikulatoren: Sprechwerkzeuge; die Organe des Hals-, Rachen- und Mundraumes, die bewegt werden müssen; um Sprachlaute zu erzeugen: Glottis; Velum; Zunge; Lippen und Unterkiefer.

Assimilation: Annäherung eines Lautes an einen Nachbarlaut bei der Aussprache durch Anpassung der Sprechbewegungen; z.B. wird das 'N' in 'unklar' aufgrund der Assimilation zum 'K' wie 'NG' ausgesprochen.

Content word: -> Sinnwort

Diphtong: eine Folge von zwei Vokalen, die derselben Silbe angehören. Im Deutschen gibt es die Diphtonge 'AU'; 'EI' und 'EU'.

Einzelwort: Im Sinne der vorliegenden Arbeit eine nur aus Buchstaben und Apostrophen bestehende Folge von Zeichen, die durch Wort-Trennungszeichen (Leerzeichen, Zeilen-

ende, Interpunktionen) von anderen Wörtern getrennt ist.

Elision: Löschen von Lauten beim Übergang von phonematischer zu phonetischer Beschreibung von Sprache. Beispiel: abhängig von der Sprechgeschwindigkeit kann 'morgen' auch wie 'morgn' ausgesprochen werden.

Flexion oder Flexionsendung: eine durch Deklination, Konjugation oder Komparation erzeugte Wortendung. Im Rahmen dieser Arbeit werden unter Flexionen auch stammfinale Zeichenfolgen oder Nachsilben verstanden, die mit Flexionen identisch sind, z.B. '-er' in 'Mutter' oder 'Reiter'.

Formant: Als Formanten werden die Resonanzfrequenzen des akustischen Filters 'Vokaltrakt' bezeichnet. Alle Vokale und die meisten Konsonanten sind durch charakteristische Formantlagen bzw. Formantverläufe gekennzeichnet.

Formwort: ein Wort, das im Gegensatz zum sog. Sinnwort keine lexikalische Bedeutung trägt, sondern nur strukturelle (grammatikalische) Funktion hat. Dazu gehören Artikel, Pronomen, Modalverben, Hilfsverben, Präpositionen, Konjunktionen, Partikel.

Fuge: Ein Buchstabe, der eingefügt wird, wenn mehrere Morphe-me zu Wörtern zusammengesetzt werden. Werden Stämme oder Teilwörter verbunden, kann wie z.B. in "Volk-s-tanz" nur ein 'S' als Fuge auftreten. Beim Zusammenfügen von Stämmen und Suffixen wie z.B. bei "Wand-l-ung" können auch andere Fugen wie 'L', 'R' und 'N' auftreten.

Funktionswort, function word: -> Formwort

Graphem: Ein Graphem ist die Codierung eines Phonems durch Buchstaben in orthografischem Text; z.B. ist 'SCH' die Codierung von 'ʃ' im Wort 'Schule'.

Grundkluster: sind alle -> Kluster; die in deutschen Morphemen auftreten können.

häufige Ausnahmen: hochfrequente Wörter der deutschen Sprache, die nicht den allgemeinen Transkriptionsregeln und Akzentuierungsregeln folgen.

Intonation: Mit Intonation im engeren Sinne werden die Veränderungen der Grundfrequenz bezeichnet, die in jeder menschlichen Äußerung auftreten. Im weiteren Sinn entspricht der Intonation die Änderung der Parameter Grundfrequenz, Lautdauer und Lautintensität.

Kluster: eine Folge von zusammenhängenden Konsonanten oder Vokalgraphemen. Z.B. zerfällt das Wort 'Wort' in:

W - initiales Konsonantenkluster

O - mediales Vokalkluster

RT - finales Konsonantenkluster

Koartikulation: Ineinandergreifen der einzelnen Artikulationsbewegungen im Zusammenhang der gesprochenen Rede, was zu Modifikationen der Einzellaute führt.

Kompositum: Eine Wortzusammensetzung, die mehr als einen Stamm enthält.

Konsonant: Konsonanten sind die Laute der Sprache, die nicht allein, sondern nur zusammen mit Vokalen ausgesprochen werden können (Mitlaute). Sie lassen sich klassifizieren nach dem Artikulationsort, d.h. der Stelle im Mund- und Rachenraum, an der die größte Verengung bzw. der Verschuß erfolgt, nach dem Grad der Verengung, nach der Art der Anregung und nach der Nasalierung. Im Deutschen unterscheidet man:

stimmlose Plosive: 'p', 't', 'k'

stimmhafte Plosive: 'b', 'd', 'g'

stimmlose Frikative: 'f', 's', 'ʃ', 'ç', 'x', 'h'

stimmhafte Frikative: 'v', 'z', 'ʒ', 'j', 'r'

Nasale: 'm', 'n', 'ɳ'

Lateral: 'l'

Konstituente: Syntaktische "Baugruppe" eines Satzes. Der Satz "Der Mann geht heute in den Garten" zerfällt z.B. in die Konstituenten "der Mann", "geht", "heute" und "in den Garten". Die letzten drei Konstituenten wiederum lassen sich zu einer übergeordneten Konstituente, der sog. Verbalphrase zusammenfassen.

Morphem: Kleinste bedeutungstragende Einheit der Sprache. Es gibt verschiedene Arten von Morphemen: Stämme, Vorsilben, Nachsilben und Endungen. Das Wort "unvergeßliche" zerfällt z.B. in:

un	:	Relationsmorphem oder Präfix oder Vorsilbe
ver	:	Relationsmorphem oder Präfix oder Vorsilbe
gess	:	Grundmorphem oder Stamm
lich	:	Relationsmorphem oder Suffix oder Nachsilbe
e	:	Flexionsmorphem oder Flexion oder Endung

Morphologie: "Formenlehre". Wie die Syntax Regeln angibt zum Aufbau von Sätzen aus Wörtern, gibt die Morphologie Regeln an zum Aufbau von Wörtern aus Morphemen.

Opposition: Gegenüberstellung von Lauten, die in nur einem bedeutungsunterscheidenden Merkmal verschieden sind. Z.B. gibt es eine 'stimmhaft'-'stimmlos'-Opposition für die Plosivlaute 'b': 'p', 'd': 't', 'g': 'k'.

Parser: Ein Parser ist ein Programm zur syntaktischen Analyse für (natürliche und formale) Sprachen.

Partikel: Alle Wörter, die nicht den flektierbaren Wortklassen angehören.

Phonem: kleinste bedeutungsunterscheidende sprachliche Einheit; z.B. 'r' : 'f' in 'rein' : 'fein'. Phoneme können durch

verschiedene -> Allophone realisiert werden; z.B. 'r' in initialer und finaler Position in 'Ruder'. Phoneme sind sprachabhängig; z.B. ist der Knacklaut im Deutschen im Gegensatz zu anderen Sprachen kein Phonem, weil nicht bedeutungsunterscheidend. Dies führt zu unterschiedlichen Phoneminventaren für verschiedene Sprachen. Man unterscheidet zwischen Vokalphonemen (-> Vokal) und Konsonantenphonemen (-> Konsonant).

phonematisch - phonetisch: Eine phonetische Beschreibung der Sprache geht von gesprochener Sprache aus und codiert sie in Form von Lauten/Allophonen unter Beachtung von Koartikulationen; Elisionen und Assimilationen. Eine phonematische Beschreibung geht von orthografischem Text aus und codiert ihn in Form von Phonemen. Eine phonematische Beschreibung kann in unterschiedliche phonetische Beschreibungen überführt werden.

Phonemisierung: Umsetzung von Text in eine (morphembezogene) phonematische Lautschrift.

Phonetik: Die Lehre von den Phonemen und ihrer (konkreten) Realisierung beim Sprechakt.

Phonologie: Die Lehre von den (abstrakten) Phonemen und ihren Beziehungen untereinander. Die Phonologie baut auf der Phonetik auf.

Phrase: Eine phonetische Phrase ist eine durch Intonation und Pausen markierte Einheit der Rede; eine syntaktische Phrase ist ein aus einem oder mehreren Wörtern bestehendes Satzglied.

Präfix: Vorsilbe. Im streng linguistischen Sinne sind Präfixe Wortbestandteile, die vor dem Stamm auftreten und nicht allein stehen können, z.B. 'ge-'; 'ent-' und 'ver-'. Im Rahmen dieser Arbeit werden unter Präfixen auch Präpositionen verstanden, die vor dem Stamm auftreten, z.B. 'vor-' in 'Vorsilbe'.

Prosodie: Prosodie im engeren Sinne bezeichnet laut Duden den Satzrhythmus; im weiteren Sinn ist der Verlauf der drei Parameter Grundfrequenz, Lautdauer und Lautstärke in einer Äußerung gemeint.

Schwa: sog. Murrel-'E', wie es z.B. in 'dunkle' realisiert wird. Der Status des Schwa ist umstritten. Es kann sowohl als Phonem als auch als allophonische Variante zum kurzen 'E' klassifiziert werden.

Semantik: Ebene der Sprachforschung; auf der die Beziehungen von sprachlichen Einheiten zu den von ihnen bezeichneten Inhalten erforscht werden.

Sinnwort: ein Wort, das eine lexikalische Bedeutung trägt. Dazu gehören Substantive, Adjektive, Numeralien, Adverbien und Verben.

Suffix: Nachsilbe. Im streng linguistischen Sinn sind Suffixe Wortbestandteile, die hinter dem Stamm auftreten und nicht allein stehen können; z.B. '-lich' und '-ig'. Im Rahmen dieser Arbeit werden unter Suffixen auch bestimmte finale Teile von fremdstämmigen Stämmen verstanden, z.B. '-iv' in 'Motiv'.

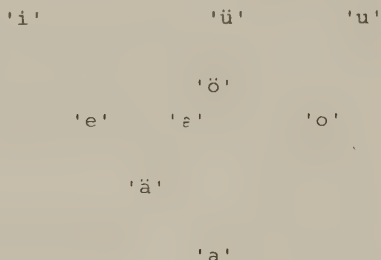
Syntax: Lehre von der Funktion der Wortarten im Satz. Die Syntax beschreibt die Regularitäten des Satzes, insbesondere seine konstruktiven Beziehungen.

Transkription: Umsetzung von geschriebener oder gesprochener Sprache in Lautschrift.

Umlaut: in der vorliegenden Arbeit einer der drei Vokale 'Ä', 'Ö' oder 'Ü'.

Vokal: Vokale sind die Laute der Sprache; die silbischen Charakter haben können; d.h. Vokale können, für sich allein stehend, eine Silbe bilden, während -> Konsonan-

ten nur vor oder nach dem Vokal im Silbenkern auftreten. Vokale lassen sich nach dem Grad der Mundöffnung und der Lage des Zungengipfels klassifizieren. Man erhält dann für die deutsche Sprache ein Vokaldreieck:



In dieser Darstellung entspricht die X-Achse in etwa dem Grad der Mundöffnung bei der Aussprache; d.h. die rechts stehenden Vokale werden als geschlossen; die links stehenden werden als offen bezeichnet. Die Y-Achse entspricht der Lage des Zungengipfels als Artikulationsort beim Sprechen, d.h. ein 'i' wird vorne, ein 'u' wird hinten artikuliert. Artikulationsort und Grad der Mundöffnung geben Hinweise auf die Lage des ersten und zweiten -> Formanten. Bis auf das Schwa können alle Vokale sowohl als kurze als auch als lange Vokale auftreten. Lange und kurze Vokale haben leicht unterschiedliche Formantlagen und in betonter Position unterschiedliche -> Vokaldauern. Neben den im Vokaldreieck dargestellten Vokalen gibt es im Deutschen noch die drei Diphthonge 'ei'; 'eu' und 'au'; die als Kombinationen der o.a. Vokale dargestellt werden können; und einige nasalierte Vokale; die nur in Fremdwörtern auftreten.

Vokaldauer: Als Vokaldauer wird im Rahmen dieser Arbeit die physikalisch meßbare Dauer von Vokalen bezeichnet, im Gegensatz zur -> Vokallänge, die eine qualitative phonologische Größe ist. Die Vokaldauer ist nicht nur vom Laut selbst; sondern auch vom Lautumfeld; der Position im Satz und von der Betonung abhängig.

Vokallänge: Im Gegensatz zur Vokaldauer ist die Vokallänge keine quantitative, sondern eine qualitative phonologische Größe. Lange Vokale haben nicht grundsätzlich, sondern nur in betonten Positionen eine meßbar größere Dauer als kurze Vokale.

Anhang 2: Initiale und finale Grundkluster

2.1 Liste der initialen Grundkluster

Kluster	Häufigkeit	Kluster	Häufigkeit
BL	9	TH	1
BR	9	TR	9
CH	9	WR	3
DR	9	ZW	7
DW	1		
FL	9		
FR	9		
GL	9		
GN	5		
GR	9		
KL	9		
KN	7		
KR	9		
PF	7		
PFL	5		
PFN	1		
PFR	3		
PL	7		
PR	7		
RH	1		
SCH	9		
SCHL	9		
SCHM	7		
SCHN	7		
SCHR	7		
SCHW	9		
SP	9		
SPL	5		
SPR	7		
ST	9		
STR	8		

2.2 Liste der finalen Grundkluster

Kluster	Häufigkeit	Kluster	Häufigkeit
BB	4	LN	1
BD	1	LP	5
BS	1	LS	4
BST	1	LSCH	3
BT	1	LST	3
CH	9	LT	9
CH	9	LZ	5
CHB	1	MB	1
CHS	7	MD	3
CHT	9	MK	1
CK	9	MM	9
CKT	1	MP	7
DD	5	MPF	7
DT	1	MPS	3
FF	8	MS	5
FT	7	MST	1
GD	1	MT	3
GG	5	NCH	3
GT	1	ND	9
HL	9	NF	3
HM	9	NFT	1
HN	8	NG	9
HR	9	NGST	1
KS	1	NK	9
KT	1	NN	9
LB	7	NS	7
LCH	4	NSCH	4
LD	7	NST	5
LF	7	NT	9
LFT	1	NZ	7
LG	7	PF	9
LK	7	PP	9
LL	9	PS	5
LM	5	PT	1

Kluster	Häufigkeit
RP	9
RSS	1
RSS7	1
RCH	6
RCH7	1
RD	7
RF	7
RF7	1
RG	7
RK	7
RKS	1
RKT	1
RL	5
RM	7
RN	9
RMS7	1
RNT	1
RP	5
RP7	3
RS	7
RS	3
RSCH	5
RS7	7
RT	9
RD	7
SCB	9
SP	7
SS	9
ST	9
TM	1
TSCH	7
TT	9
TD	9
TTT	1
WV	1
XT	3

Anhang 3: Falsche Zerlegungen durch die Klusteranalyse

Von der Klusteranalyse nicht erkannte oder falsch zerlegte Wortzusammensetzungen unter den 55/220 mehrfach auftretenden Zusammensetzungen.

Zum Darstellungsformat: Die Zahl hinter dem Wort gibt an, wie oft das Wort im Text aufgetreten ist, danach werden die synthetisierte Lautschrift sowie die korrekte Lautschrift für den unterstrichenen Teil des Wortes angegeben. Der restliche Teil des Wortes wird korrekt umgesetzt bzw. nicht von der Morphemgrenze beeinflusst. Die Lautschrift wird in Anhang 5 beschrieben. Wörter, die durch die fehlerhafte Zerlegung schwer verständlich werden, sind mit einem 'x' gekennzeichnet; unverständliche Wörter mit einem doppelten 'x'.

3.1. Nicht erkannte Morphemgrenzen

3.1.1 Das erste Wort endet auf 'e'

Frage <u>b</u> ogen	28	E	/	\$
Hilfe <u>l</u> eistung	11	E	/	\$
Frage <u>s</u> tellung	10	EST	/	\$\$/T
Pflege <u>d</u> ienst	19	E	/	\$
Helle <u>f</u> eld	6	E	/	\$
xx Pflege <u>g</u> everein	3	EWER	/	\$fER?
Eingabe <u>t</u> ext	3	E	/	\$
Ausgabe <u>s</u> ystem	2	E	/	\$
gemeinde <u>n</u> ah	2	E	/	\$

3.1.2 Das erste Wort endet auf '-en'

Zusammen <u>e</u> fassung	3	E	/	\$
Zusammen <u>a</u> rbeit	2	ENA	/	\$N?A
xx Woche <u>n</u> ende	2	\$N\$	/	\$N?E
Wochen <u>s</u> tunden	2	ENST	/	\$NS/T
Inne <u>n</u> dienst	2	E	/	\$

Bogen <u>u</u> ntersuchung	18	ENUNTERZ	/	\$N?UNT\$RZ
Bogen <u>e</u> rgebnisse	2	ENER	/	\$N?ER
Daten <u>a</u> nalyse	2	ENA	/	\$N?A
Sonnen <u>s</u> chein	2	E	/	\$

3.1.3 Das erste Wort endet auf '-er'

Wörter <u>b</u> uch	3	E	/	\$
xx Schwer <u>w</u> erpunkt	3	E	/	e
xx Schwer <u>b</u> ehinderten	3	ERBE	✓	eRB\$
Ander <u>e</u> rseits	3	ERER	/	\$R\$R
Aller <u>d</u> ings	3	E	/	\$
Träger <u>o</u> rganisation	2	ERO	/	\$R?O
Sonder <u>d</u> asein	2	E	/	\$
hier <u>d</u> urch	2	.		
hierzu <u>l</u> ande	2			

3.1.4 Das erste Wort ist fremdstämmig

xx Ziv <u>i</u> ldienst	4	I	/	i
xx Literatur <u>u</u> bericht	2	URBE	/	uRB\$
xx Pauschal <u>a</u> bezuschung	2	ALBE	/	aLB\$

3.1.5 Sonstige Zusammensetzungen

Satz <u>i</u> ntonation	6	TZI	/	TZ?I
zeit <u>i</u> ntensiv	6	TI	/	T?I
Sprach <u>s</u> ynthese	3			
xx Schreib <u>a</u> rbeiten	2	BA	/	P?A
fünf <u>u</u> ndzwanzig	2	FU	/	F?U
Antwort <u>a</u> lternativen	2	TA	/	T?A

3.2. Nicht korrekt bestimmte Morphemgrenzen

3.2.1 Falsche Fuge '-s-' bei 'st-':

voll <u>s</u> tändig	3	ST	/	S/T
Verselb <u>s</u> tändigung	3	ST	/	S/T

Personal <u>str</u> uktur	3	ST	/	S/T
Beratungs <u>st</u> ellen	3	ST	/	S/T
Antrag <u>s</u> teller	2	ST	/	S/T
Werk <u>st</u> ätten	2	ST	/	S/T
Lehr <u>st</u> uhl	2	ST	/	S/T

3.2.2 Sonstige falsche Fugen -'s'

Sprachau- <u>s</u> -gabe	4	A/S	/	?A/S
Mar- <u>s</u> -berg	2	aRS	/	ARS

3.2.3 Nicht erkanntes Fugen -'s'

x Untersuchung- <u>s</u> ergebnisse	2	ZER	/	S?ER
x Untersuchung- <u>s</u> intensität	2	ZI	/	S?I
x Hilf- <u>s</u> organisation	4	ZO	/	S?O
x Betreuung- <u>s</u> angebot	2	ZAN/GE	/	S?AN/G\$
x Betreuung- <u>s</u> aufgaben	2	ZA/	/	S?A/

3.2.4 Falsche Grenzbestimmung ohne Fuge

xx Nominal <u>ph</u> -rasen	3	ALFr	/	aLFr
x Fran- <u>k</u> reich	3	aN	/	AN/
x Stüt- <u>z</u> wörter	4	yT	/	YT
Mikrop- <u>r</u> ozessor	2			
x Antwor- <u>t</u> rücklauf	2	oR	/	OR

COMP Format: Befehl, 1. zu vergl. Z-kette, 2. zu vergl.
 Z-kette, ... ; ende-kennzeichnung

Verändernde Vergleichsbefehle:

CADD Format: Befehl, 1. zu vergl. Z-kette, Lautschrift zu 1.,
2. zu vergl. Z-kette, Lautschr. zu 2.; ... , ende

VCADD1 : Vergleiche 1; 2 bzw. 3 Zeichen lange Zeichenketten,
VCADD2 : die vor oder nach der zu verifizierenden Zeichenfolge
VCADD3 : stehen. Bei erfolgreichem Vergleich werden die zu-
NCADD1 : sätzlichen Zeichen dem zu verifizierenden Morphem
NCADD2 : zugeschlagen und mit der angegebenen Lautschrift
NCADD3 : versehen. Die Lautschrift der ursprünglichen Zeichen-
kette wird nicht verändert.

CNPH Format: Befehl, 1. zu vergl. Z-kette; 1. Lautschrift,
2. zu vergl. Z-kette, 2. Lautschrift, ... ; ende

```
VCNPH1 : Vergleiche 1, 2 bzw. 3 Zeichen lange Zeichenketten,  
VCNPH2 : die vor oder nach der zu verifizierenden Zeichenfolge  
VCNPH3 : stehen. Bei erfolgreichem Vergleich werden die zu-  
NCNPH1 : sätzlichen Zeichen dem zu verifizierenden Morphem  
NCNPH2 : zugeschlagen und das gesamte Morphem mit einer neuen  
NCNPH3 : Lautschrift versehen.
```

CNBI Format: Befehl, neue Betonungsinformation, 1. zu vergl.
 Z-kette, 2. zu vergl. Z-kette, ... , ende

```
VCNBI1 :   Vergleiche 1, 2 bzw. 3 Zeichen lange Zeichenketten,
VCNBI2 :   die vor oder nach der zu verifizierenden Zeichenfolge
VCNBI3 :   stehen. Bei erfolgreichem Vergleich wird nur die
NCNBI1 :   Betonungsinformation geändert.
NCNBI2 :
NCNBI3 :
```

NCNPO Format: Befehl, neue Lautschrift; 1. zu vergl. Z-kette,
2. zu vergl. Z-kette, ... , ende

NCNPO1 : Vergleiche 1, 2, bzw. 3 Zeichen lange Zeichenketten,
NCNPO2 : die nach der zu verifizierenden Zeichenfolge stehen.
NCNPO3 : Bei erfolgreichem Vergleich wird die Lautschrift des
Morphems geändert.

Weitere Befehle:

XPHNWB Format: Befehl, neue Lautschrift; ende

Die Lautschrift des Morphems wird geändert, falls zuvor die Betonungsinformation geändert wurde.

CTLKL Format: Befehl

VCTLKL : Prüfe, ob vor bzw. nach dem zu verifizierenden Morphem
NCTLKL : ein initiales bzw. finales Grundcluster folgt. Wenn
 nein, Verifizierung abbrechen.

NOOPER Format: Befehl

 Mache gar nichts. Dieser Befehl wird nur zu Testzwecken benötigt.

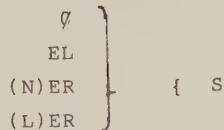
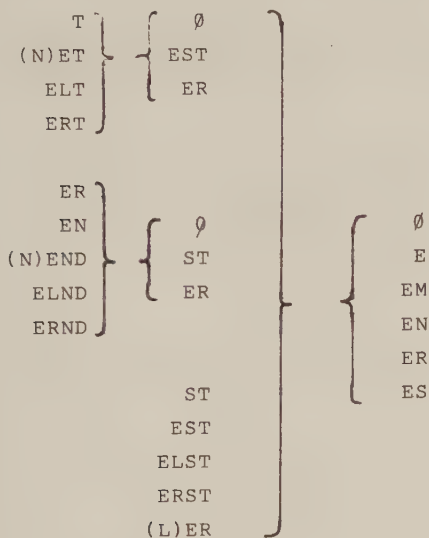
ENDE

Format: Befehl

Dieser Befehl kennzeichnet das Ende des Verifizierungseintrags und führt zum Abbruch der Verifizierung.

Anhang 5: Zur Abtrennung von Endungen

5.1 Grammatik zur Abtrennung von Endungen



Die Schreibung {} soll bedeuten, daß jede Buchstabenfolge vor der Klammer mit jeder Zeichenfolge nach der Klammer kombiniert werden kann. Das Zeichen 'Ø' bedeutet die leere Menge (kein Zeichen). In runde Klammern gesetzte Zeichen sind Fugen, die vor einer Endung stehen können und mit berücksichtigt werden.

5.2 Regeln zur Verifizierung der Verkleinerungssilbe '-chen'

1. Steht vor '-chen' kein 'S', ist '-chen' als Endung zu markieren.
2. Steht vor '-chen' ein 'S', ist '-chen' dann eine Endung, wenn das letzte Vokalzeichen davor ein 'Ä', ein 'Ö' oder ein 'Ü' ist.
3. Steht vor '-chen' ein 'S' und sind die letzten Vokalzeichen davor 'ÄU', liegen nur Endungen vor wie bei den Verkleinerungsformen von 'Haus', 'Klaus', 'Laus' und 'Maus'.

5.3 Regeln zur Verifizierung von mit 'T' beginnenden Endungen

1. Steht vor 'T' ein Vokal, kann 'T' nicht Anfang einer Endung sein.
2. Liegt vor 'T' eine von der Klusteranalyse markierte Wortgrenze, ist 'T' Anfang einer Endung.
3. Steht vor 'T' ein 'L', 'F', 'T' oder 'D'; ist 'T' kein Endungsanfang.
4. Stehen vor 'T' andere Konsonanten als 'K', 'N' oder 'R', ist 'T' Endungsanfang.
5. Steht 'ER' vor 'T'; ist 'T' Endungsanfang.
6. Steht vor 'KT', 'NT' oder 'RT' noch ein weiterer Konsonant, ist 'T' Anfang einer Endung, sonst nicht.

Anhang 6: Die verwendete Lautschrift

Für alle Lautschriftzeichen sind in der folgenden Tabelle IPA-Notation (IPA), eigene Lautschrift (Ausdruck), programminterne Bezeichnung (Label) und interne Codierung (Code) angegeben.

IPA	Ausdruck	Label	Code(hex)	Beispielwort
y:	y	LUEWRT	F1	H <u>ü</u> te
Y	Y	UEWORT	F9	H <u>ü</u> tte
ø:	q	LOEWRT	F2	l <u>ö</u> sen
œ	Q	OEWORT	FA	l <u>ö</u> schen
æ:	v	LAEWRT	F3	w <u>ä</u> hlen
ɛ	V	AEWRT	FB	W <u>ä</u> lle
u:	u	LUWORT	F4	M <u>u</u> t
U	U	KUWORT	FC	M <u>u</u> tter
o:	o	LOWORT	F5	P <u>o</u> l
O	O	KOWORT	FD	P <u>o</u> llen
i:	i	LIWORT	F6	M <u>i</u> ete
I	I	KIWORT	FE	M <u>i</u> tte
e:	e	LEWORT	F7	z <u>e</u> hn
ɛ	E	KEWORT	FF	Z <u>e</u> ntner
a:	a	LAWORT	F8	W <u>a</u> hl
a	A	KAWORT	00	W <u>a</u> ll
ə	§	SCHWA	02	G <u>e</u> hör
ao	A/	AUWORT	EE	P <u>a</u> use
oe	O/	EUWORT	EF	h <u>e</u> ute
ae	E/	EIWORT	F0	P <u>f</u> eil
b	B	BWORT	01	B <u>a</u> nd
p	P	PWORT	0F	P <u>a</u> rk
d	D	DWORT	03	d <u>o</u> rt
t	T	TWORT	13	T <u>o</u> rf
g	G	GWORT	06	G <u>a</u> be
k	K	KWORT	0A	K <u>a</u> bel

IPA	Ausdruck	Label	Code(hex)	Beispielwort
m	M	MWORT	0C	<u>M</u> ann
n	N	NWORT	0D	<u>N</u> ame
ŋ	N/	NGWORT	0E	s <u>an</u> g
l	L	LWORT	0B	<u>L</u> ied
r	r	RWORT	11	<u>R</u> uf
ʁ	R	ERWORT	1A	U <u>f</u> er
f	F	FWORT	05	<u>v</u> ier
v	W	WWORT	16	<u>w</u> ir
s	S	SWORT	12	<u>L</u> os
z	Z	WSWORT	14	<u>L</u> ose
ʃ	S/	SHWORT	08	<u>S</u> chule
ʒ	Z/	GGWORT	15	Gar <u>ag</u> e
ç	C	ICHWRT	04	<u>L</u> icht
j	J	JWORT	09	<u>J</u> agd
x	X	ACHWRT	10	<u>N</u> acht
h	H	HWORT	07	<u>H</u> und
ʔ	?	GLSTOP	18	Be <u>am</u> ter
ks	K/	XWORT	17	! <u>M</u> ax
ts	T/	ZWORT	19	! <u>Z</u> elt
	_/	LAP AUS	EB	
	_	PAUSE	EC	
	'	APOST	ED	!

Die mit einem '!' gekennzeichneten Lautschriftzeichen dürfen nicht an der Treiber-Schnittstelle übergeben werden.

Anhang 7: Zur Textvorverarbeitung

7.1 Die Aussprache von Sonderzeichen

Sonderzeichen, die immer ausgesprochen werden:

#	-	PFUND
\$	-	DOLLAR
%	-	PROZENT
&	-	UND
(	-	RUNDE KLAMMER AUF
)	-	RUNDE KLAMMER ZU
*	-	STERN
+	-	PLUS
/	-	SCHRÄGSTRICH
<	-	KLEINER
=	-	GLEICH
>	-	GRÖßER
§	-	PARAGRAPH
^	-	AUFWÄRTSPFEIL
_	-	UNTERSTREICHUNG
'	-	AKZENT

Sonderzeichen, die nur im amerikanischen ASCII-Zeichensatz ausgesprochen werden:

[	-	ECKIGE KLAMMER AUF
\	-	BACKSLASH
]	-	ECKIGE KLAMMER ZU
{	-	GESCHWEIFTE KLAMMER AUF
	-	SENKRECHTER STRICH
}	-	GESCHWEIFTE KLAMMER ZU
~	-	CIRKA

7.2 Von der Textvorverarbeitung verwendete Steuerzeichen

Code(hex)	Bezeichnung	Verwendung
08	BS	letztes Zeichen löschen
18	EM	letzte Zeile löschen
01	SOH	letzte Zeile wiederholen
1A	SUB	Software-Reset = Neuinitialisierung
1B	ESC	leitet 'Escape-Sequenz' zur Übergabe von Parametern ein

Aufbau einer 'Escape-Sequenz':

1. Zeichen: 'ESC'
2. Zeichen: Codierung des Parameters
3. + 4. Zeichen: neuer Wert des Parameters; hexadezimal, jedes Halb-Byte durch ein ASCII-Zeichen codiert.

Es können die folgenden Parameter codiert werden:

Code	Parameter
T	Sprechgeschwindigkeit
S	Zeitlimit für die Eingabe
Z	Zeichensatz
L	Lautstärke
F	Stimmelage

LEBENS LAUF:

- 1951 Geboren an 15.4.1951 in Voerde/Nörrrh. als Sohn von
Werkzeugmacher Wilhelm Isselhorst und Margarete
Isselhorst, geb. Tebbe.
- 1957 - 1962 Besuch der Volksschule in Voerde.
- 1962 - 1967 Besuch der Kreisrealschule in Voerde.
- 1967 - 1970 Besuch des städtischen neusprachlichen Theodor-
Heuss-Gymnasiums mit Aufbauzweig für Realschul-
absolventen in Dinslaken, Abitur im Juni 1970.
- 1969 Adoptiert von Elektromeister Gerhard Rühl und Helene
Rühl, geb. Isselhorst.
- 1970 - 1972 Freiwilliger bei der Luftwaffe der Bundeswehr.
- 1972 - 1978 Studium der Elektrotechnik an der Ruhr-Universität
Bochum, Diplom-Hauptprüfung im April 1978.
- 1977 Heirat mit Annemarie Neukäter im Dezember 1977.
- 1978 Wissenschaftliche Hilfskraft am Lehrstuhl für
Allgemeine Elektrotechnik und Akustik der Ruhr-
Universität Bochum von Juni bis Dezember.
- 1979 Wissenschaftlicher Angestellter am o.a. Lehrstuhl
von Januar bis September.
- seit 1979 Wissenschaftlicher Assistent am o.a. Lehrstuhl seit
Oktober 1979.

